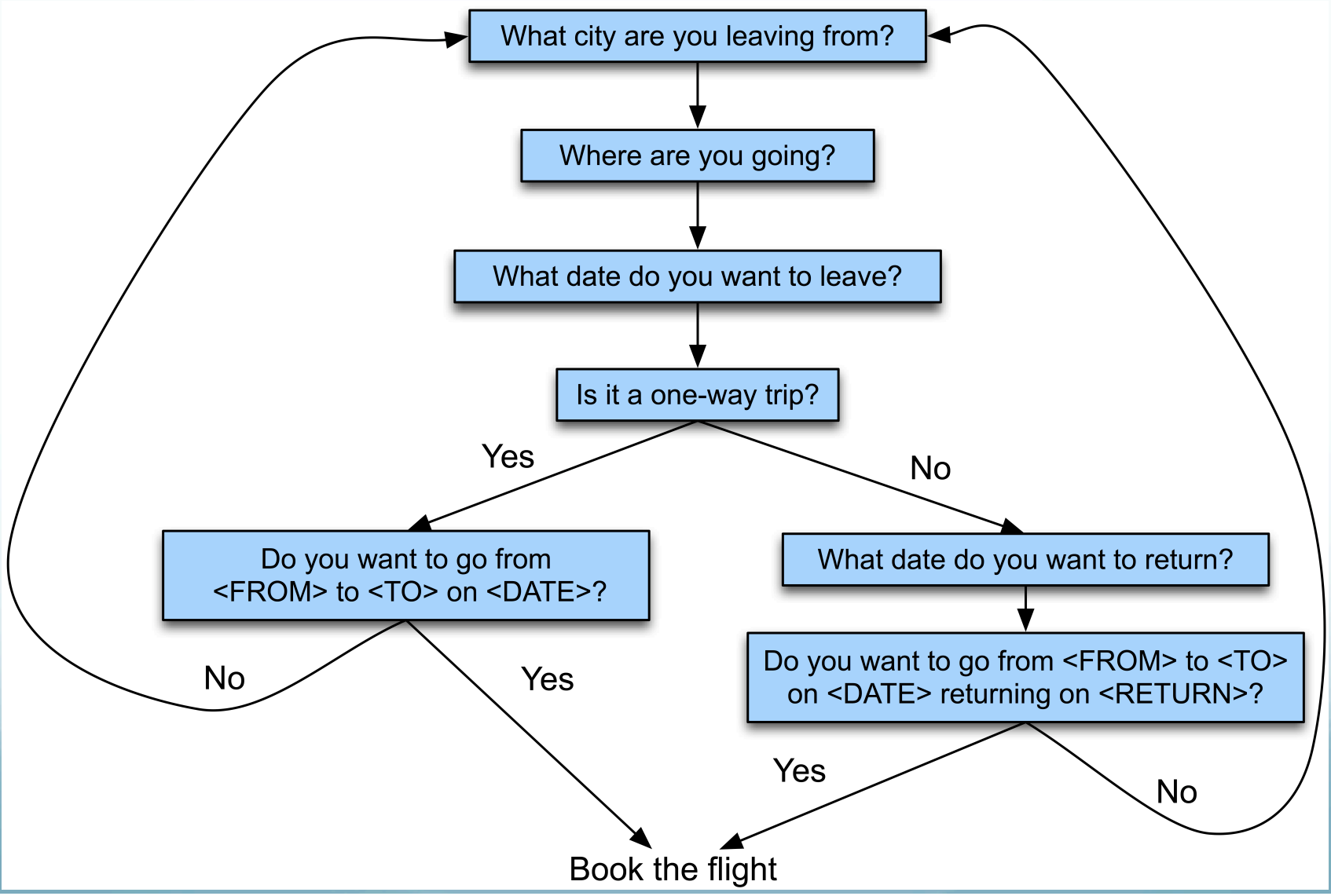
Dialogue Management

Ling575
Discourse and Dialogue
May 18, 2011

Dialog Management Types

- Finite-State Dialog Management
- Frame-based Dialog Management
  - Initiative
  - VoiceXML
  - Design and evaluation
- Information State Management
  - Dialogue Acts
    - Recognition & generation
- Statistical Dialogue Managemant (POMDPs)

Finite-State Management



Pros and Cons

- Advantages
  - Straightforward to encode
  - Clear mapping of interaction to model
  - Well-suited to simple information access
  - System initiative
- Disadvantages
  - Limited flexibility of interaction
    - Constrained input – single item
    - Fully system controlled
    - Restrictive dialogue structure, order
  - Ill-suited to complex problem-solving

Frame-based Dialogue Management

- Finite-state too limited, stilted, irritating
- More flexible dialogue

Frame-based Dialogue Management

- Essentially form-filling
  - User can include any/all of the pieces of form
  - System must determine which entered, remain

Frame-based Dialogue Management

- Essentially form-filling
  - User can include any/all of the pieces of form
  - System must determine which entered, remain

Slot	Question
ORIGIN CITY	"From what city are you leaving?"
DESTINATION CITY	"Where are you going?"
DEPARTURE TIME	"When would you like to leave?"
ARRIVAL TIME	"When do you want to arrive?"

Frame-based Dialogue Management

- Essentially form-filling
  - User can include any/all of the pieces of form
  - System must determine which entered, remain

Slot	Question
ORIGIN CITY	"From what city are you leaving?"
DESTINATION CITY	"Where are you going?"
DEPARTURE TIME	"When would you like to leave?"
ARRIVAL TIME	"When do you want to arrive?"

- System may have multiple frames
  - E.g. flights vs restrictions vs car vs hotel
  - Rules determine next action, question, information presentation

Frames and Initiative

- Mixed initiative systems:
  - A) User/System can shift control arbitrarily, any time
    - Difficult to achieve

Frames and Initiative

- Mixed initiative systems:
  - A) User/System can shift control arbitrarily, any time
    - Difficult to achieve
  - B) Mix of control based on prompt type
- Prompts:

Frames and Initiative

- Mixed initiative systems:
  - A) User/System can shift control arbitrarily, any time
    - Difficult to achieve
  - B) Mix of control based on prompt type
- Prompts:
  - Open prompt:

Frames and Initiative

- Mixed initiative systems:
  - A) User/System can shift control arbitrarily, any time
    - Difficult to achieve
  - B) Mix of control based on prompt type
- Prompts:
  - Open prompt: ‘How may I help you?’
    - Open-ended, user can respond in any way
  - Directive prompt:

Frames and Initiative

- Mixed initiative systems:
  - A) User/System can shift control arbitrarily, any time
    - Difficult to achieve
  - B) Mix of control based on prompt type
- Prompts:
  - Open prompt: ‘How may I help you?’
    - Open-ended, user can respond in any way
  - Directive prompt: ‘Say yes to accept call, or no o.w.’
    - Stipulates user response type, form

Initiative, Prompts, Grammar

- Prompt type tied to active grammar
  - System must recognize suitable input
    - Restrictive vs open-ended

Initiative, Prompts, Grammar

- Prompt type tied to active grammar
  - System must recognize suitable input
    - Restrictive vs open-ended
- Shift from restrictive to open
  - Tune to user: Novice vs Expert

Initiative, Prompts, Grammar

- Prompt type tied to active grammar
  - System must recognize suitable input
    - Restrictive vs open-ended
- Shift from restrictive to open
  - Tune to user: Novice vs Expert

Grammar	Prompt Type: Open	Prompt Type: Directive
Restrictive	*Doesn't make sense*	System Initiative
Non-Restrictive	User Initiative	Mixed Initiative

Figure 24.10 Operational definition of initiative, following Singh et al. (2002).

Dialogue Management: Confirmation

- Miscommunication common in SDS
  - "Error spirals" of sequential errors
    - Highly problematic
  - Recognition, recovery crucial
- Confirmation strategies can detect, mitigate
  - Explicit confirmation:

Dialogue Management: Confirmation

- Miscommunication common in SDS
  - "Error spirals" of sequential errors
    - Highly problematic
  - Recognition, recovery crucial
- Confirmation strategies can detect, mitigate
  - Explicit confirmation:
    - Ask for verification of each input
  - Implicit confirmation:

Dialogue Management: Confirmation

- Miscommunication common in SDS
  - "Error spirals" of sequential errors
    - Highly problematic
  - Recognition, recovery crucial
- Confirmation strategies can detect, mitigate
  - Explicit confirmation:
    - Ask for verification of each input
  - Implicit confirmation:
    - Include input information in subsequent prompt

Confirmation Strategies

- Explicit:

S:	**Which city do you want to leave from?**
U:	Baltimore.
S:	**Do you want to leave from Baltimore?**
U:	Yes.
U:	I'd like to fly from Denver Colorado to New York City on September twenty first in the morning on United Airlines
S:	**Let's see then. I have you going from Denver Colorado to New York on September twenty first. Is that correct?**
U:	Yes

Confirmation Strategy

- Implicit:

U:	I want to travel to Berlin
S:	**When do you want to travel to Berlin?**
U2:	Hi I'd like to fly to Seattle Tuesday Morning
A3:	**Traveling to Seattle on Tuesday, August eleventh in the morning. Your full name?**

Pros and Cons

- Grounding of user input
  - Weakest grounding
    - I.e. continued att'n, next relevant contibution

Pros and Cons

- Grounding of user input
  - Weakest grounding insufficient
    - I.e. continued att'n, next relevant contibution
  - Explicit:

Pros and Cons

- Grounding of user input
  - Weakest grounding insufficient
    - I.e. continued att'n, next relevant contibution
  - Explicit: highest: repetition
  - Implicit:

Pros and Cons

- Grounding of user input
  - Weakest grounding insufficient
    - I.e. continued att'n, next relevant contibution
  - Explicit: highest: repetition
  - Implicit: demonstration, display
- Explicit;

Pros and Cons

- Grounding of user input
  - Weakest grounding insufficient
    - I.e. continued att'n, next relevant contibution
  - Explicit: highest: repetition
  - Implicit: demonstration, display
- Explicit;
  - Pro: easier to correct; Con: verbose, awkward, non-human
- Implicit:

Pros and Cons

- Grounding of user input
  - Weakest grounding insufficient
    - I.e. continued att'n, next relevant contibution
  - Explicit: highest: repetition
  - Implicit: demonstration, display
- Explicit;
  - Pro: easier to correct; Con: verbose, awkward, non-human
- Implicit:
  - Pro: more natural, efficient; Con: less easy to correct

Rejection

- System recognition confidence is too low
- System needs to reprompt
  - Often repeatedly

Rejection

- System recognition confidence is too low
- System needs to reprompt
  - Often repeatedly
    - Out-of-vocabulary, out-of-grammar inputs
- Strategies: Progressive prompting

Rejection

- System recognition confidence is too low
- System needs to reprompt
  - Often repeatedly
    - Out-of-vocabulary, out-of-grammar inputs
- Strategies: Progressive prompting
  - Initially: ‘rapid reprompting’: ‘What?’, ‘Sorry?’

Rejection

- System recognition confidence is too low
- System needs to reprompt
  - Often repeatedly
    - Out-of-vocabulary, out-of-grammar inputs
- Strategies: Progressive prompting
  - Initially: ‘rapid reprompting’: ‘What?’, ‘Sorry?’
  - Later: increasing detail

- Progressive prompting

System:	When would you like to leave?
Caller:	Well, um, I need to be in New York in time for the first World Series game.
System:	<reject>. Sorry, I didn't get that. Please say the month and day you'd like to leave.
Caller:	I wanna go on October fifteenth.

VoiceXML

- W3C standard for simple frame-based dialogues
  - Fairly common in commercial settings
- Construct forms, menus
  - Forms get field data
    - Using attached prompts
    - With specified grammar (CFG)
    - With simple semantic attachments

Simple VoiceXML Example

```
<form>
  <field name="transporttype">
     <prompt>
        Please choose airline, hotel, or rental car.
     </prompt>
     <grammar type="application/x=nuance-gsl">
       [airline hotel "rental car"]
     </grammar>
  </field>
  <block>
    <prompt>
     You have chosen <value expr="transporttype">.
    </prompt>
  </block>
</form>
```

Frame-based Systems: Pros and Cons

- Advantages

Frame-based Systems: Pros and Cons

- Advantages
  - Relatively flexible input – multiple inputs, orders
  - Well-suited to complex information access (air)
  - Supports different types of initiative
- Disadvantages

Frame-based Systems: Pros and Cons

- Advantages
  - Relatively flexible input – multiple inputs, orders
  - Well-suited to complex information access (air)
  - Supports different types of initiative
- Disadvantages
  - Ill-suited to more complex problem-solving
    - Form-filling applications

Dialogue Manager Tradeoffs

- Flexibility vs Simplicity/Predictability
  - System vs User vs Mixed Initiative
  - Order of dialogue interaction
  - Conversational “naturalness” vs Accuracy
  - Cost of model construction, generalization, learning, etc

Dialog Systems Design

- User-centered design approach:
  - Study user and task:
    - Interview users; record human-human interactions; systems

Dialog Systems Design

- User-centered design approach:
  - Study user and task:
    - Interview users; record human-human interactions; systems
  - Build simulations and prototypes:
    - Wizard-of-Oz systems (WOZ): Human replaces system
      - Can assess issues in partial system; simulate errors, etc

Dialog Systems Design

- User-centered design approach:
  - Study user and task:
    - Interview users; record human-human interactions; systems
  - Build simulations and prototypes:
    - Wizard-of-Oz systems (WOZ): Human replaces system
      - Can assess issues in partial system; simulate errors, etc
  - Iteratively test on users:
    - Redesign prompts (email subdialog)
    - Identify need for barge-in

SDS Evaluation

- Goal: Determine overall user satisfaction
  - Highlight systems problems; help tune
- Classically: Conduct user surveys

TTS Performance	Was the system easy to understand ?
ASR Performance	Did the system understand what you said?
Task Ease	Was it easy to find the message/flight/train you wanted?
Interaction Pace	Was the pace of interaction with the system appropriate?
User Expertise	Did you know what you could say at each point?
System Response	How often was the system sluggish and slow to reply to you?
Expected Behavior	Did the system work the way you expected it to?
Future Use	Do you think you'd use the system in the future?

Figure 24.14 User satisfaction survey, adapted from Walker et al. (2001).

SDS Evaluation

- User evaluation issues:

SDS Evaluation

- User evaluation issues:
  - Expensive; often unrealistic; hard to get real user to do
- Create model correlated with human satisfaction
- Criteria:

SDS Evaluation

- User evaluation issues:
  - Expensive; often unrealistic; hard to get real user to do
- Create model correlated with human satisfaction
- Criteria:
  - Maximize task success
    - Measure task completion: % subgoals; Kappa of frame values

SDS Evaluation

- User evaluation issues:
  - Expensive; often unrealistic; hard to get real user to do
- Create model correlated with human satisfaction
- Criteria:
  - Maximize task success
    - Measure task completion: % subgoals; Kappa of frame values
  - Minimize task costs
    - Efficiency costs: time elapsed; # turns; # error correction turns
    - Quality costs: # rejections; # barge-in; concept error rate

PARADISE Model

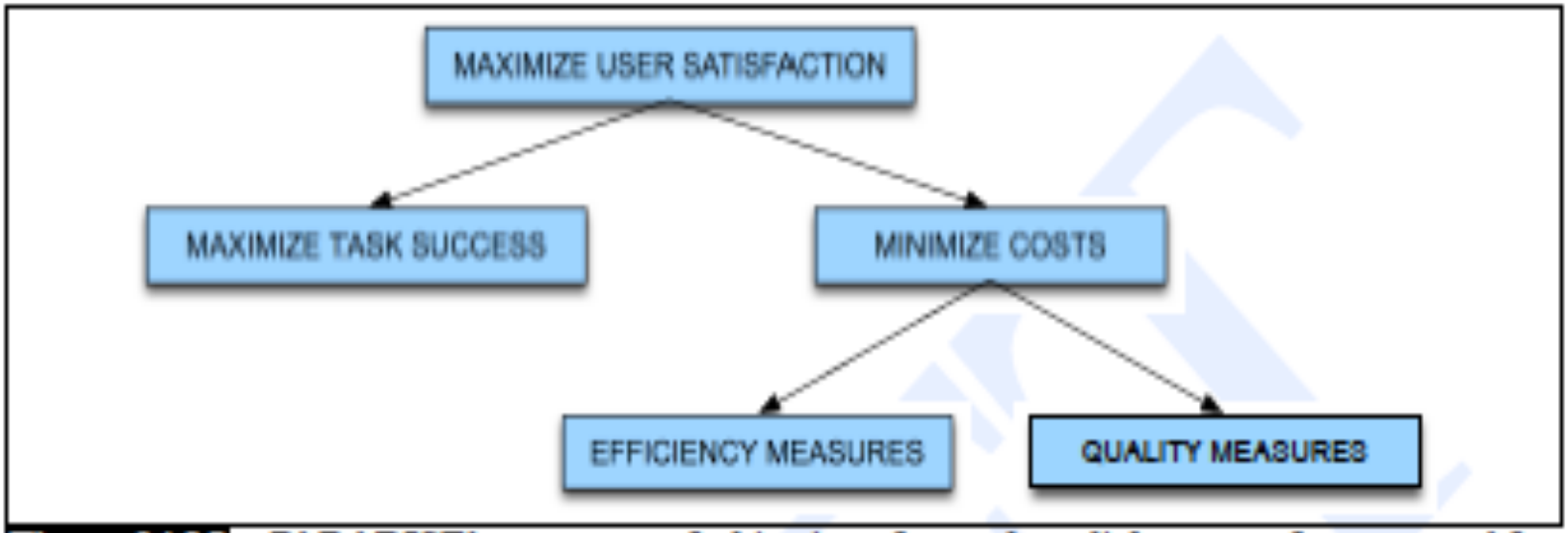


Figure 24.15 PARADISE's structure of objectives for spoken dialogue performance. After Walker et al. (1997).

PARADISE Model

- Compute user satisfaction with questionnaires

PARADISE Model

- Compute user satisfaction with questionnaires
- Extract task success and costs measures from corresponding dialogs
  - Automatically or manually

PARADISE Model

- Compute user satisfaction with questionnaires
- Extract task success and costs measures from corresponding dialogs
  - Automatically or manually
- Perform multiple regression:
  - Assign weights to all factors of contribution to Usat
  - Task success, Concept accuracy key

PARADISE Model

- Compute user satisfaction with questionnaires
- Extract task success and costs measures from corresponding dialogs
  - Automatically or manually
- Perform multiple regression:
  - Assign weights to all factors of contribution to Usat
  - Task success, Concept accuracy key
- Allows prediction of accuracy on new dialog w/Q&A

Information State Dialogue Management

- Problem: Not every task is equivalent to form-filling
- Real tasks require:

Information State Dialogue Management

- Problem: Not every task is equivalent to form-filling
- Real tasks require:
  - Proposing ideas, refinement, rejection, grounding, clarification, elaboration, etc

Information State Dialogue Management

- Problem: Not every task is equivalent to form-filling
- Real tasks require:
  - Proposing ideas, refinement, rejection, grounding, clarification, elaboration, etc
- Information state models include:
  - Information state
  - Dialogue act interpreter
  - Dialogue act generator
  - Update rules
  - Control structure

Information State Systems

- Information state :
  - Discourse context, grounding state, intentions, plans.

Information State Systems

- Information state :
  - Discourse context, grounding state, intentions, plans.
- Dialogue acts:
  - Extension of speech acts, to include grounding acts
    - Request-inform; Confirmation

Information State Systems

- Information state :
  - Discourse context, grounding state, intentions, plans.
- Dialogue acts:
  - Extension of speech acts, to include grounding acts
    - Request-inform; Confirmation
- Update rules
  - Modify information state based on DAs

Information State Systems

- Information state :
  - Discourse context, grounding state, intentions, plans.
- Dialogue acts:
  - Extension of speech acts, to include grounding acts
    - Request-inform; Confirmation
- Update rules
  - Modify information state based on DAs
    - When a question is asked

Information State Systems

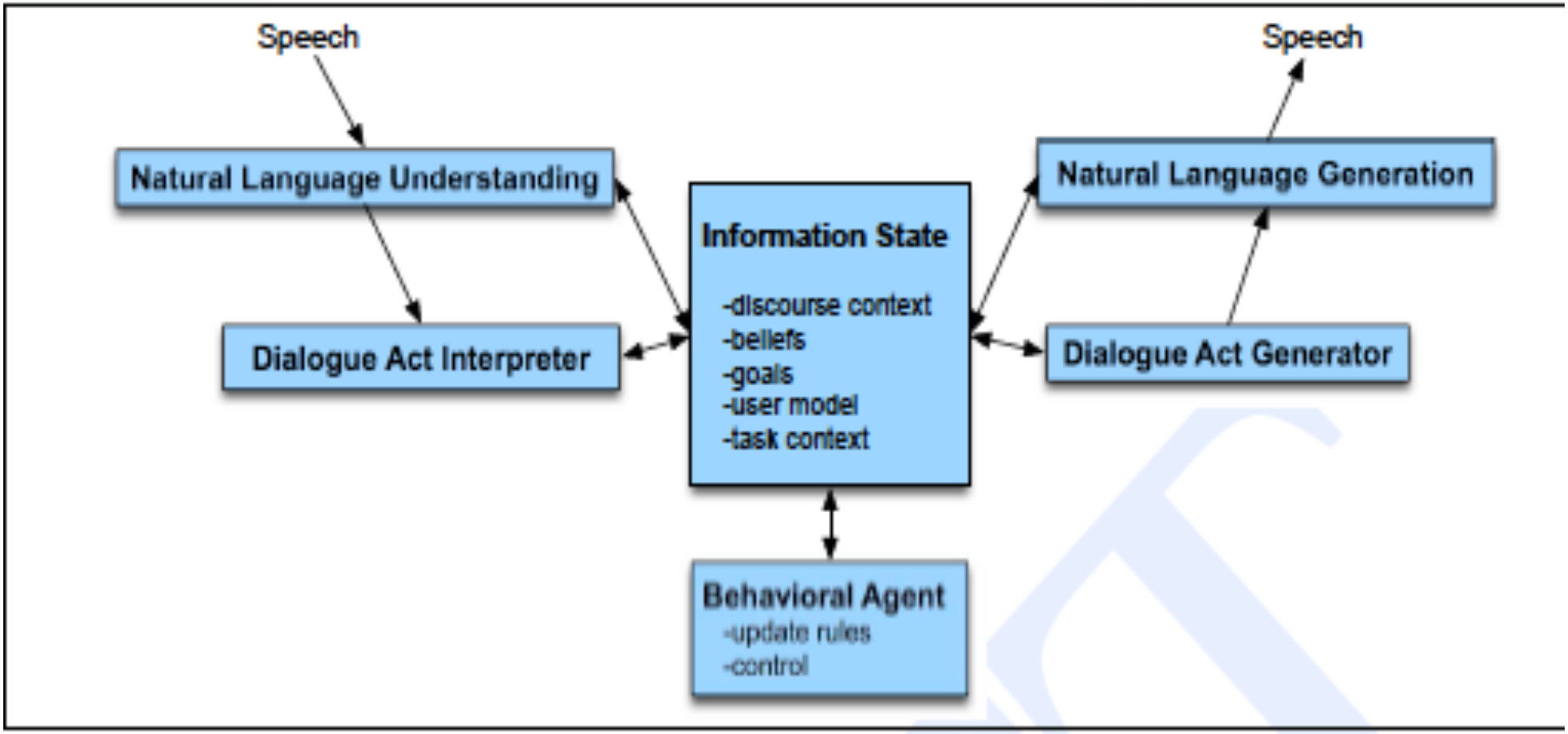
- Information state :
  - Discourse context, grounding state, intentions, plans.
- Dialogue acts:
  - Extension of speech acts, to include grounding acts
    - Request-inform; Confirmation
- Update rules
  - Modify information state based on DAs
    - When a question is asked, answer it
    - When an assertion is made,

Information State Systems

- Information state :
  - Discourse context, grounding state, intentions, plans.
- Dialogue acts:
  - Extension of speech acts, to include grounding acts
    - Request-inform; Confirmation
- Update rules
  - Modify information state based on DAs
    - When a question is asked, answer it
    - When an assertion is made,
      - Add information to context, grounding state

Information State Architecture

- Simple ideas, complex execution



Dialogue Acts

- Extension of speech acts
  - Adds structure related to conversational phenomena
    - Grounding, adjacency pairs, etc

Dialogue Acts

- Extension of speech acts
  - Adds structure related to conversational phenomena
    - Grounding, adjacency pairs, etc
- Many proposed tagsets
  - Verbmobil: acts specific to meeting sched domain

Dialogue Acts

- Extension of speech acts
  - Adds structure related to conversational phenomena
    - Grounding, adjacency pairs, etc
- Many proposed tagsets
  - Verbmobil: acts specific to meeting sched domain
  - DAMSL: Dialogue Act Markup in Several Layers
    - Forward looking functions: speech acts
    - Backward looking function: grounding, answering

Dialogue Acts

- Extension of speech acts
  - Adds structure related to conversational phenomena
    - Grounding, adjacency pairs, etc
- Many proposed tagsets
  - Verbmobil: acts specific to meeting sched domain
  - DAMSL: Dialogue Act Markup in Several Layers
    - Forward looking functions: speech acts
    - Backward looking function: grounding, answering
  - Conversation acts:
    - Add turn-taking and argumentation relations

Verbmobil DA

- 18 high level tags

Tag	Example
THANK	*Thanks*
GREET	*Hello Dan*
INTRODUCE	*It's me again*
BYE	*Allright bye*
REQUEST-COMMENT	*How does that look?*
SUGGEST	*from thirteenth through seventeenth June*
REJECT	*No Friday I'm booked all day*
ACCEPT	*Saturday sounds fine,*
REQUEST-SUGGEST	*What is a good day of the week for you?*
INIT	*I wanted to make an appointment with you*
GIVE_REASON	*Because I have meetings all afternoon*
FEEDBACK	*Okay*
DELIBERATE	*Let me check my calendar here*
CONFIRM	*Okay, that would be wonderful*
CLARIFY	*Okay, do you mean Tuesday the 23rd?*
DIGRESS	*[we could meet for lunch] and eat lots of ice cream*
MOTIVATE	*We should go to visit our subsidiary in Munich*
GARBAGE	*Oops, I-*

Figure 24.17 The 18 high-level dialogue acts used in Verbmobil-1, abstracted over a total of 43 more specific dialogue acts. Examples are from Jekat et al. (1995).

Dialogue Act Interpretation

- Automatically tag utterances in dialogue
- Some simple cases:
  - Will breakfast be served on USAir 1557?

Dialogue Act Interpretation

- Automatically tag utterances in dialogue
- Some simple cases:
  - **YES-NO-Q**: Will breakfast be served on USAir 1557?
  - I don't care about lunch.

Dialogue Act Interpretation

- Automatically tag utterances in dialogue
- Some simple cases:
  - **YES-NO-Q:** Will breakfast be served on USAir 1557?
  - **Statement:** I don't care about lunch.
  - Show be flights from L.A. to Orlando

Dialogue Act Interpretation

- Automatically tag utterances in dialogue
- Some simple cases:
  - **YES-NO-Q:** Will breakfast be served on USAir 1557?
  - **Statement:** I don't care about lunch.
  - **Command:** Show be flights from L.A. to Orlando
- Is it always that easy?
  - Can you give me the flights from Atlanta to Boston?

Dialogue Act Interpretation

- Automatically tag utterances in dialogue
- Some simple cases:
  - **YES-NO-Q:** Will breakfast be served on USAir 1557?
  - **Statement:** I don't care about lunch.
  - **Command:** Show be flights from L.A. to Orlando
- Is it always that easy?
  - Can you give me the flights from Atlanta to Boston?
    - Syntactic form: question; Act: request/command
  - Yeah.

Dialogue Act Interpretation

- Automatically tag utterances in dialogue
- Some simple cases:
  - **YES-NO-Q:** Will breakfast be served on USAir 1557?
  - **Statement:** I don't care about lunch.
  - **Command:** Show be flights from L.A. to Orlando
- Is it always that easy?
  - Can you give me the flights from Atlanta to Boston?
  - Yeah.
    - Depends on context: Y/N answer; agreement; back-channel

Dialogue Act Ambiguity

- Indirect speech acts

A		I was wanting to make some arrangements for a trip that I'm going to be taking uh to LA uh beginning of the week after next.
B		OK uh let me pull up your profile and I'll be right with you here. [pause]
B		**And you said you wanted to travel next week?**
A		Uh yes.

Dialogue Act Ambiguity

- Indirect speech acts

A	OPEN-OPTION	I was wanting to make some arrangements for a trip that I'm going to be taking uh to LA uh beginning of the week after next.
B		OK uh let me pull up your profile and I'll be right with you here. [pause]
B		**And you said you wanted to travel next week?**
A		Uh yes.

Dialogue Act Ambiguity

- Indirect speech acts

A	OPEN-OPTION	I was wanting to make some arrangements for a trip that I'm going to be taking uh to LA uh beginning of the week after next.
B	HOLD	OK uh let me pull up your profile and I'll be right with you here. [pause]
B		**And you said you wanted to travel next week?**
A		Uh yes.

Dialogue Act Ambiguity

- Indirect speech acts

A	OPEN-OPTION	I was wanting to make some arrangements for a trip that I'm going to be taking uh to LA uh beginning of the week after next.
B	HOLD	OK uh let me pull up your profile and I'll be right with you here. [pause]
B	CHECK	**And you said you wanted to travel next week?**
A		Uh yes.

Dialogue Act Ambiguity

- Indirect speech acts

A	OPEN-OPTION	I was wanting to make some arrangements for a trip that I'm going to be taking uh to LA uh beginning of the week after next.
B	HOLD	OK uh let me pull up your profile and I'll be right with you here. [pause]
B	CHECK	**And you said you wanted to travel next week?**
A	ACCEPT	Uh yes.

Dialogue Act Recognition

- How can we classify dialogue acts?
- Sources of information:

Dialogue Act Recognition

- How can we classify dialogue acts?
- Sources of information:
  - Word information:
    - *Please, would you*: request; *are you*: yes-no question

Dialogue Act Recognition

- How can we classify dialogue acts?
- Sources of information:
  - Word information:
    - *Please, would you*: request; *are you*: yes-no question
    - N-gram grammars
  - Prosody:

Dialogue Act Recognition

- How can we classify dialogue acts?
- Sources of information:
  - Word information:
    - *Please, would you*: request; *are you*: yes-no question
    - N-gram grammars
  - Prosody:
    - Final rising pitch: question; final lowering: statement
    - Reduced intensity: *Yeah:* agreement vs backchannel

Dialogue Act Recognition

- How can we classify dialogue acts?
- Sources of information:
  - Word information:
    - *Please, would you*: request; *are you*: yes-no question
    - N-gram grammars
  - Prosody:
    - Final rising pitch: question; final lowering: statement
    - Reduced intensity: *Yeah:* agreement vs backchannel
  - Adjacency pairs:

Dialogue Act Recognition

- How can we classify dialogue acts?
- Sources of information:
  - Word information:
    - *Please, would you*: request; *are you*: yes-no question
    - N-gram grammars
  - Prosody:
    - Final rising pitch: question; final lowering: statement
    - Reduced intensity: *Yeah:* agreement vs backchannel
  - Adjacency pairs:
    - Y/N question, agreement vs Y/N question, backchannel
    - DA bi-grams

Task & Corpus

- Goal:
  - Identify dialogue acts in conversational speech

Task & Corpus

- Goal:
  - Identify dialogue acts in conversational speech
- Spoken corpus: Switchboard
  - Telephone conversations between strangers
  - Not task oriented; topics suggested
  - 1000s of conversations
    - recorded, transcribed, segmented

Dialogue Act Tagset

- Cover general conversational dialogue acts
  - No particular task/domain constraints

Dialogue Act Tagset

- Cover general conversational dialogue acts
  - No particular task/domain constraints
- Original set: ~50 tags
  - Augmented with flags for task, conv mgmt
    - 220 tags in labeling: some rare

Dialogue Act Tagset

- Cover general conversational dialogue acts
  - No particular task/domain constraints
- Original set: ~50 tags
  - Augmented with flags for task, conv mgmt
    - 220 tags in labeling: some rare
- Final set: 42 tags, mutually exclusive
  - SWBD-DAMSL
  - Agreement: K=0.80 (high)

Dialogue Act Tagset

- Cover general conversational dialogue acts
  - No particular task/domain constraints
- Original set: ~50 tags
  - Augmented with flags for task, conv mgmt
    - 220 tags in labeling: some rare
- Final set: 42 tags, mutually exclusive
  - SWBD-DAMSL
  - Agreement: K=0.80 (high)
- 1,155 conv labeled: split into train/test

Common Tags

- **Statement** & **Opinion**: declarative +/- op
- **Question**: Yes/No&Declarative: form, force
- **Backchannel:** Continuers like uh-huh, yeah
- **Turn Exit/Adandon:** break off, +/- pass
- **Answer :** Yes/No, follow questions
- **Agreement**: Accept/Reject/Maybe

Probabilistic Dialogue Models

- HMM dialogue models

Probabilistic Dialogue Models

- HMM dialogue models
  - States = Dialogue acts; Observations: Utterances
    - Assume decomposable by utterance
    - Evidence from true words, ASR words, prosody

$$d^* = \underset{d}{\operatorname{argmax}} P(d \mid o) = \underset{d}{\operatorname{argmax}} \frac{P(o \mid d)P(d)}{P(o)} = \underset{d}{\operatorname{argmax}} P(o \mid d)P(d)$$

Probabilistic Dialogue Models

- HMM dialogue models
  - States = Dialogue acts; Observations: Utterances
    - Assume decomposable by utterance
    - Evidence from true words, ASR words, prosody

$$d^* = \underset{d}{\operatorname{argmax}} P(d \mid o) = \underset{d}{\operatorname{argmax}} \frac{P(o \mid d)P(d)}{P(o)} = \underset{d}{\operatorname{argmax}} P(o \mid d)P(d)$$

$$P(o \mid d) = P(f \mid d)P(W \mid d)$$

Probabilistic Dialogue Models

- HMM dialogue models
  - States = Dialogue acts; Observations: Utterances
    - Assume decomposable by utterance
    - Evidence from true words, ASR words, prosody

$$d^* = \underset{d}{\operatorname{argmax}} P(d \mid o) = \underset{d}{\operatorname{argmax}} \frac{P(o \mid d)P(d)}{P(o)} = \underset{d}{\operatorname{argmax}} P(o \mid d)P(d)$$

$$P(o \mid d) = P(f \mid d)P(W \mid d)$$

$$P(W \mid d) = \prod_{i=2}^{N} P(w_i \mid w_{i-1}, w_{i-2} \ldots w_{i-N+1}, d)$$

Probabilistic Dialogue Models

- HMM dialogue models
  - States = Dialogue acts; Observations: Utterances
    - Assume decomposable by utterance
    - Evidence from true words, ASR words, prosody

$$d^* = \underset{d}{\operatorname{argmax}} P(d \mid o) = \underset{d}{\operatorname{argmax}} \frac{P(o \mid d)P(d)}{P(o)} = \underset{d}{\operatorname{argmax}} P(o \mid d)P(d)$$

$$P(o \mid d) = P(f \mid d)P(W \mid d)$$

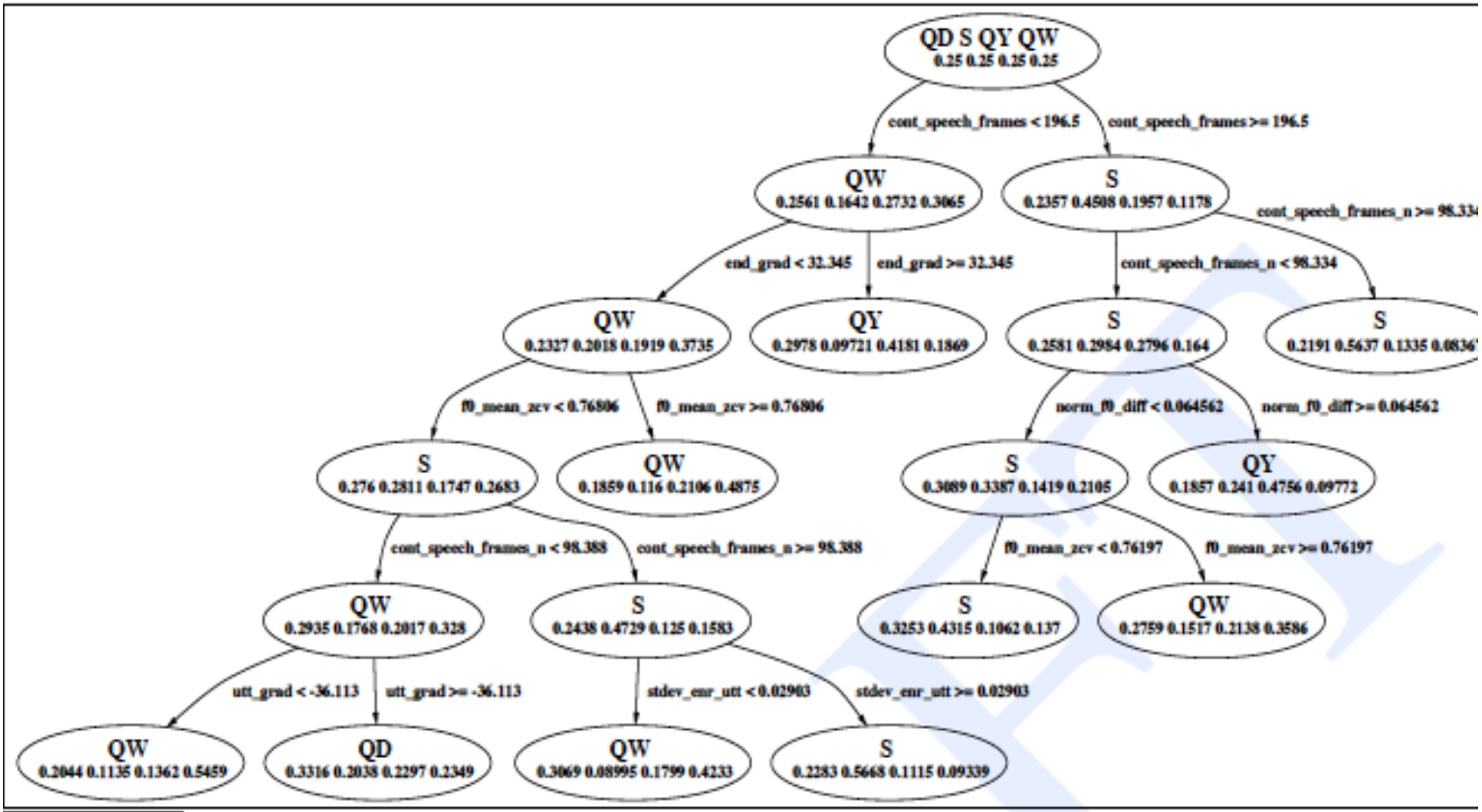
$$P(W \mid d) = \prod_{i=2}^{N} P(w_i \mid w_{i-1}, w_{i-2} \ldots w_{i-N+1}, d)$$

$$d^* = \underset{d}{\operatorname{argmax}} P(d \mid d_{t-1})P(f \mid d)P(W \mid d)$$

DA Classification - Prosody

- Features:
  - Duration, pause, pitch, energy, rate, gender
    - Pitch accent, tone
- Results:
  - Decision trees: 5 common classes
    - 45.4% - baseline=16.6%

Prosodic Decision Tree



DA Classification -Words

- Words
  - Combines notion of discourse markers and collocations:
    - e.g. uh-huh=Backchannel
  - Contrast: true words, ASR 1-best, ASR n-best
- Results:
  - Best: 71%- true words, 65% ASR 1-best

DA Classification - All

- Combine word and prosodic information
  - Consider case with ASR words and acoustics

DA Classification - All

- Combine word and prosodic information
  - Consider case with ASR words and acoustics
  - Prosody classified by decision trees
    - Incorporate decision tree posteriors in model for P(f|d)

DA Classification - All

- Combine word and prosodic information
  - Consider case with ASR words and acoustics
  - Prosody classified by decision trees
    - Incorporate decision tree posteriors in model for P(f|d)

$$d^* = P(d \mid d_{t-1}) \frac{P(d \mid f)}{P(d)} \prod_{i=2}^{N} P(w_i \mid w_{i-1} \ldots w_{i-N+1}, d)$$

- Slightly better than raw ASR

Integrated Classification

- Focused analysis
  - Prosodically disambiguated classes
    - Statement/Question-Y/N and Agreement/Backchannel
    - Prosodic decision trees for agreement vs backchannel
      - Disambiguated by duration and loudness

Integrated Classification

- Focused analysis
  - Prosodically disambiguated classes
    - Statement/Question-Y/N and Agreement/Backchannel
    - Prosodic decision trees for agreement vs backchannel
      - Disambiguated by duration and loudness
  - Substantial improvement for prosody+words
    - True words: S/Q: 85.9%-> 87.6; A/B: 81.0%->84.7

Integrated Classification

- Focused analysis
  - Prosodically disambiguated classes
    - Statement/Question-Y/N and Agreement/Backchannel
    - Prosodic decision trees for agreement vs backchannel
      - Disambiguated by duration and loudness
  - Substantial improvement for prosody+words
    - True words: S/Q: 85.9%-> 87.6; A/B: 81.0%->84.7
    - ASR words: S/Q: 75.4%->79.8; A/B: 78.2%->81.7
  - More useful when recognition is iffy

Many Variants

- Maptask: (13 classes)
  - Serafin & DiEugenio 2004
    - Latent Semantic analysis on utterance vectors
    - Text only
    - Game information; No improvement for DA history

Many Variants

- Maptask: (13 classes)
  - Serafin & DiEugenio 2004
    - Latent Semantic analysis on utterance vectors
    - Text only
    - Game information; No improvement for DA history
  - Surendran & Levow 2006
    - SVMs on term n-grams, prosody
    - Posteriors incorporated in HMMs
      - Prosody, sequence modeling improves

Many Variants

- Maptask: (13 classes)
  - Serafin & DiEugenio 2004
    - Latent Semantic analysis on utterance vectors
    - Text only
    - Game information; No improvement for DA history
  - Surendran & Levow 2006
    - SVMs on term n-grams, prosody
    - Posteriors incorporated in HMMs
      - Prosody, sequence modeling improves
- MRDA: Meeting tagging: 5 broad classes

Observations

- DA classification can work on open domain
  - Exploits word model, DA context, prosody
  - Best results for prosody+words
  - Words are quite effective alone – even ASR
- Questions:

Observations

- DA classification can work on open domain
  - Exploits word model, DA context, prosody
  - Best results for prosody+words
  - Words are quite effective alone – even ASR
- Questions:
  - Whole utterance models? – more fine-grained
  - Longer structure, long term features

Detecting Correction Acts

- Miscommunication is common in SDS
  - Utterances after errors misrecognized >2x as often
    - Frequently repetition or paraphrase of original input

Detecting Correction Acts

- Miscommunication is common in SDS
  - Utterances after errors misrecognized >2x as often
    - Frequently repetition or paraphrase of original input
- Systems need to detect, correct

Detecting Correction Acts

- Miscommunication is common in SDS
  - Utterances after errors misrecognized >2x as often
    - Frequently repetition or paraphrase of original input
- Systems need to detect, correct
- Corrections are spoken differently:
  - Hyperarticulated (slower, clearer) -> lower ASR conf.

Detecting Correction Acts

- Miscommunication is common in SDS
  - Utterances after errors misrecognized >2x as often
    - Frequently repetition or paraphrase of original input
- Systems need to detect, correct
- Corrections are spoken differently:
  - Hyperarticulated (slower, clearer) -> lower ASR conf.
  - Some word cues: 'No',' I meant', swearing..

Detecting Correction Acts

- Miscommunication is common in SDS
  - Utterances after errors misrecognized >2x as often
    - Frequently repetition or paraphrase of original input
- Systems need to detect, correct
- Corrections are spoken differently:
  - Hyperarticulated (slower, clearer) -> lower ASR conf.
  - Some word cues: 'No',' I meant', swearing..
- Can train classifiers to recognize with good acc.

Generating Dialogue Acts

- Generation neglected relative to generation

Generating Dialogue Acts

- Generation neglected relative to generation
- Stent (2002) model: Conversation acts, Belief model
  - Develops update rules for content planning, e.g.
    - If user releases turn, system can do 'TAKE-TURN' act
    - If system needs to summarize, use ASSERT act

Generating Dialogue Acts

- Generation neglected relative to generation
- Stent (2002) model: Conversation acts, Belief model
  - Develops update rules for content planning, i.e.
    - If user releases turn, system can do 'TAKE-TURN' act
    - If system needs to summarize, use ASSERT act
  - Identifies turn-taking as key aspect of dialogue gen.

Cue	Turn-taking acts signaled
um	KEEP-TURN, TAKE-TURN, RELEASE-TURN
<lipsmack>, <click>, so, uh	KEEP-TURN, TAKE-TURN
you know, isn't that so	ASSIGN-TURN

Figure 24.21 Language used to perform turn-taking acts, from Stent (2002).

Generating Confirmation

- Simple systems use fixed confirmation strategy
  - Implicit or explicit

Generating Confirmation

- Simple systems use fixed confirmation strategy
  - Implicit or explicit
- More complex systems can select dynamically
  - Use information state and features to decide

Generating Confirmation

- Simple systems use fixed confirmation strategy
  - Implicit or explicit
- More complex systems can select dynamically
  - Use information state and features to decide
    - Likelihood of error:
      - Low ASR confidence score
        - If very low, can reject

Generating Confirmation

- Simple systems use fixed confirmation strategy
  - Implicit or explicit
- More complex systems can select dynamically
  - Use information state and features to decide
    - Likelihood of error:
      - Low ASR confidence score
        - If very low, can reject
      - Sentence/prosodic features: longer, initial pause, pitch range

Generating Confirmation

- Simple systems use fixed confirmation strategy
  - Implicit or explicit
- More complex systems can select dynamically
  - Use information state and features to decide
    - Likelihood of error:
      - Low ASR confidence score
        - If very low, can reject
      - Sentence/prosodic features: longer, initial pause, pitch range
    - Cost of error:

Generating Confirmation

- Simple systems use fixed confirmation strategy
  - Implicit or explicit
- More complex systems can select dynamically
  - Use information state and features to decide
    - Likelihood of error:
      - Low ASR confidence score
        - If very low, can reject
      - Sentence/prosodic features: longer, initial pause, pitch range
    - Cost of error:
      - Book a flight vs looking up information
- Markov Decision Process models more detailed