



Dialog/User modeling for the Spoken Dialog System

- Research area
- ICSLP papers
 - (1) “Talking to Machines” by Steve Young
 - (2) “AT&T HELP DESK” by G. D. Fabbrizio, etc.
 - (3) “Portable, Server side Framework for VoiceXML”, B. Carpenter, etc.
- One speech application in the near future

USC SAIL
JongHo Shin
October 24, 2002



Research Area

- Statistical approaches!!!

Design and test two agents' dialogs/user modeling using equilibrium point

■ User modeling

ex) Probability, Bayesian, or Game theoretical

■ Design Spoken Dialog Management

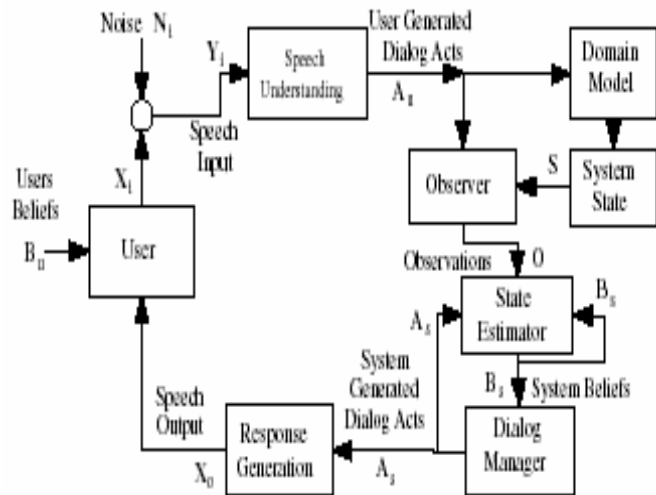
ex) Finite state model, Bayesian or Event driven

■ Spoken Dialog System Architecture

ex) DARPA Hub, Distributed, or Multi-Modal

Implement spoken dialog system for JOF(Journalism of Future) Project

Talking To Machine (Statistically Speaking) by “Steve Young” -Keynote Speaker

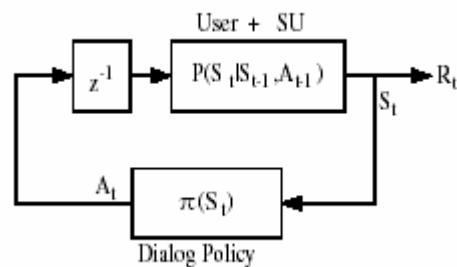


[Structure of Spoken Dialog System]

- Minimal dependence on explicit rules, and a heavy dependence on learning from data
- Optimal dialog management policy can be learned from data – POMDP
- Speech understanding component
: Estimating the posterior distribution, $P(A_u|Y)$
- Response generation
: Modeling $P(X_o|A_s)$ and sampling the distribution

Talking To Machine (Statistically Speaking)

- Dialogue management and control



- Pieraccini and Levin modelled an SDS as a fully observable MDP (Markov Decision Process)
- (Action + state) at time t \rightarrow determines the expected immediate reward, $r_{t+1} = r(S_t, A_t)$
- Goal: To find a policy which maximizes the total reward $R_t = \sum_{\tau=t}^{T-1} r_{\tau+1}$

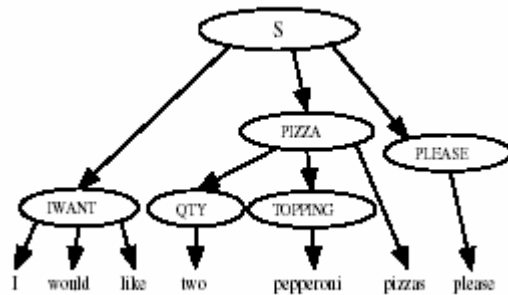
MDP Representation of a Dialogue System

<Reinforcement learning>

- (1) $V^\pi(S) = E_\pi \{R_t | S_t = S\}$: Value function
- (2) $Q^\pi(S, A) = E_\pi \{R_t | S_t = S, A_t = A\}$
- (3) $\pi^*(S) = \operatorname{argmax}_A Q^\pi(S, A)$
- (4) $Q(S, A) \leftarrow r(S, A) + \sum_{S'} P(S' | S, A) \max_{A'} Q(S', A')$: if the system transition is known
- (5) $Q(S, A) \leftarrow Q(S, A) + \alpha [r(S, A) + Q(S, A) - Q(S', A')]$: if not known, TD learning
- (6) $V^\pi(B) = \sum_S B(S) \cdot V^\pi(S)$: no info about the system state \rightarrow POMDP

Talking To Machine (Statistically Speaking)

- Semantic decoding



Example Semantic Parse Tree

- SDS depend on a semantic template grammar and some form of robust parser to extract the required semantic concepts
- Example, "I would like to two pepperoni pizzas please"

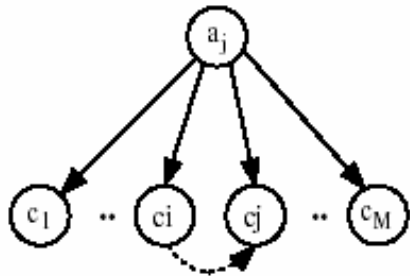
$$(1) \hat{C} = \operatorname{argmax}_C \{ P(\hat{W}|C)P(C|B_s) \} \quad C: \text{concept}, W: \text{Word}, B_s: \text{Dialogue State Belief}$$

$$(2) \approx \operatorname{argmax}_C \left\{ \prod_{t=1}^T P(w_t|w_{t-1} \dots w_{t-n+1}, c_t) P(c_t|c_{t-1} \dots c_{t-m+1}, B_s) \right\}$$

Concepts modeled as m-gram conditioned on the current system beliefs, and
The words are modeled as an n-gram conditioned on the semantic concepts

Talking To Machine (Statistically Speaking)

- Dialogue act detection



Dialog Act Detecting using a Bayesian Network

- Finite allowed dialog acts, $A=\{a_1,\dots,a_N\}$
- Finite semantic concepts, $C=\{c_1,\dots,c_M\}$
- Decision trees can be trained automatically from examples of already occurred concept sets and the corresponding dialogue acts.
- Bayesian Network allows detection of multiple dialogue acts
- $P(c_k|a_j)$ and priors, $P(a_j)$ are learned from training data, so $P(a_j|c_1,\dots,c_M)$ is computed for each act a_j



“AT&T HELP DESK”

- G.D.Fabbrizio, etc. AT&T Labs

- Objectives: Routing of calls or accessing information and handling technical problems, sales inquiries, recommendations, and troubleshooting.
- Technology extensions that are needed for speech recognition, speech synthesis, language understanding, dialog, and user interface design.
- One key issue addressed: the creation of complex services when speech data is limited or unavailable.



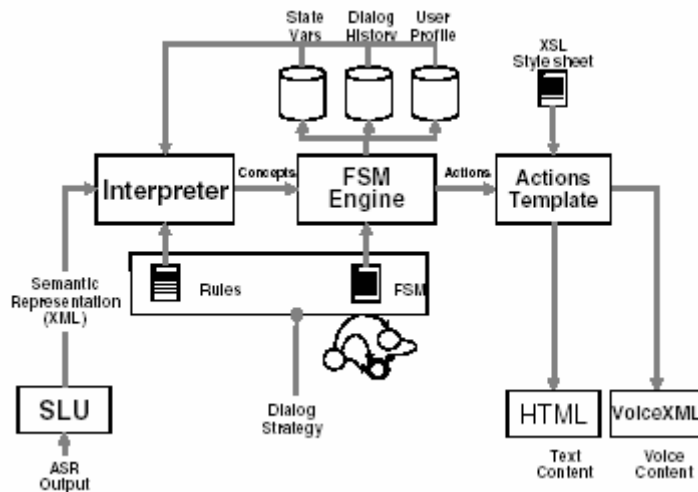
“AT&T HELP DESK”

- ASR, TTS, and Transcription and Annotation

- ASR - Initially uses a general-purpose context-dependent HMM.
 - After the system is deployed in the field, this model is adapted using Maximum a posteriori adaptation.
 - Design of stochastic language model: looked for different sources of data to achieve fast bootstrapping of language model
- TTS: uses AT&T Labs Natural Voices technology and voice fonts
 - new and customized voice fonts were created
- Transcription and annotation: mining and reusing data and models.
 - data are annotated for speech understanding purposes
 - Ex) “may I hear this again” -> tag, “discouse_repeat”

“AT&T HELP DESK”

- Dialog Management & User Interface



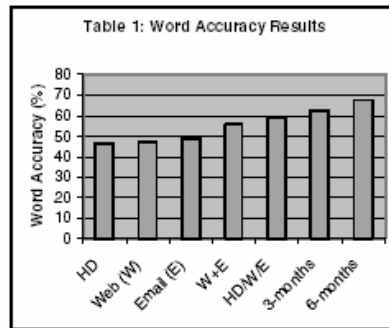
Dialog Manager Architecture

- Mixed-initiative spoken dialog
- At each dialog turn, DM accomplishes concrete task or subtask
- FSM(Finite State Machine) engine controls the actions taken in response to the interpreter output

-
- User interface design
 - Handcrafted in terms of usability and quality

“AT&T HELP DESK”

- Application & Results



- **TTS Help Desk: deployed on July 31st, 2001**
 - The service to perform routing of calls into specialized agents
- **ASR results show that 59% word accuracy without any formal data collection**
- **68% accuracy when sufficient data are available**

Table 2: Classification and task completion results

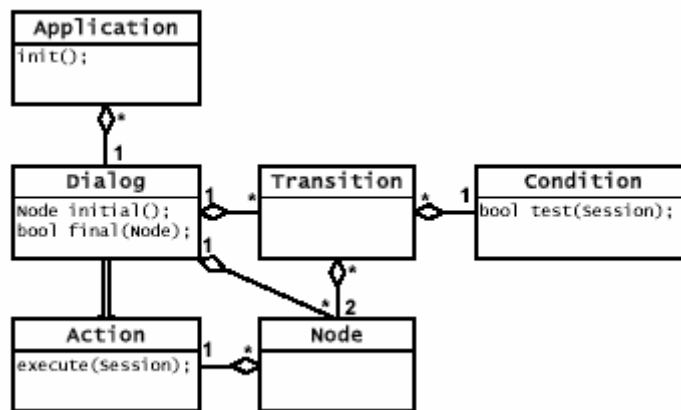
Release	V1.0	V1.1	V1.2	V1.3	V1.4
Classification	72	75	83	85	84
Routing	62	74	85	84	83
Demo	74	80	87	94	96
Information	80	78	71	68	72

Task completion

- **Consistent improvement in the classification accuracy and the task completion rates**
- **Detecting the correct semantic tag was 84% in V1.4**
- **The percentage of correct system action given an input request was 85% in average**

“A Portable, Server-Side Dialog Framework for VoiceXML

- Etude Dialog Manager from “SpeechWorks”



Dialog classes hierarchy

- The fundamental building block of the framework is the Action interface
- Support Frame format for the data retrieval and delivery in session
- Errors and Exceptions are handled in each dialog module itself
- Transitions are defined by a human
- Dialog modules are connected by transitions defined

Speech Application in the Future



Speech

Dialog is involved !!



Increase the inner temperature of the house

Turn on the light

Turn on the TV

Home Networking