

Design & Evaluation

Ling575
Spoken Dialog Systems
April 17, 2013

Designing Dialog

- Apply user-centered design

Designing Dialog

- Apply user-centered design
 - Study user and task: How?

Designing Dialog

- Apply user-centered design
 - Study user and task: How?
 - Interview potential users, record human-human tasks
 - Study how the user interacts with the system

Designing Dialog

- Apply user-centered design
 - Study user and task: How?
 - Interview potential users, recorded human-human tasks
 - Study how the user interacts with the system
 - But it's not built yet....

Designing Dialog

- Apply user-centered design
 - Study user and task: How?
 - Interview potential users, recorded human-human tasks
 - Study how the user interacts with the system
 - But it's not built yet....
 - Wizard-of-Oz systems: Simulations
 - User thinks they're interacting with a system, but it's driven by a human
 - Prototypes

Designing Dialog

- Apply user-centered design
 - Study user and task: How?
 - Interview potential users, recorded human-human tasks
 - Study how the user interacts with the system
 - But it's not built yet....
 - Wizard-of-Oz systems: Simulations
 - User thinks they're interacting with a system, but it's driven by a human
 - Prototypes
 - Iterative redesign:
 - Test system: see how users really react, what problems occur, correct, repeat

SDS Evaluation

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

SDS Evaluation

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

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
 - Minimize task costs
 - Efficiency costs: time elapsed; # turns; # error correction turns
 - Quality costs: # rejections; # barge-in; concept error rate

PARADISE Model

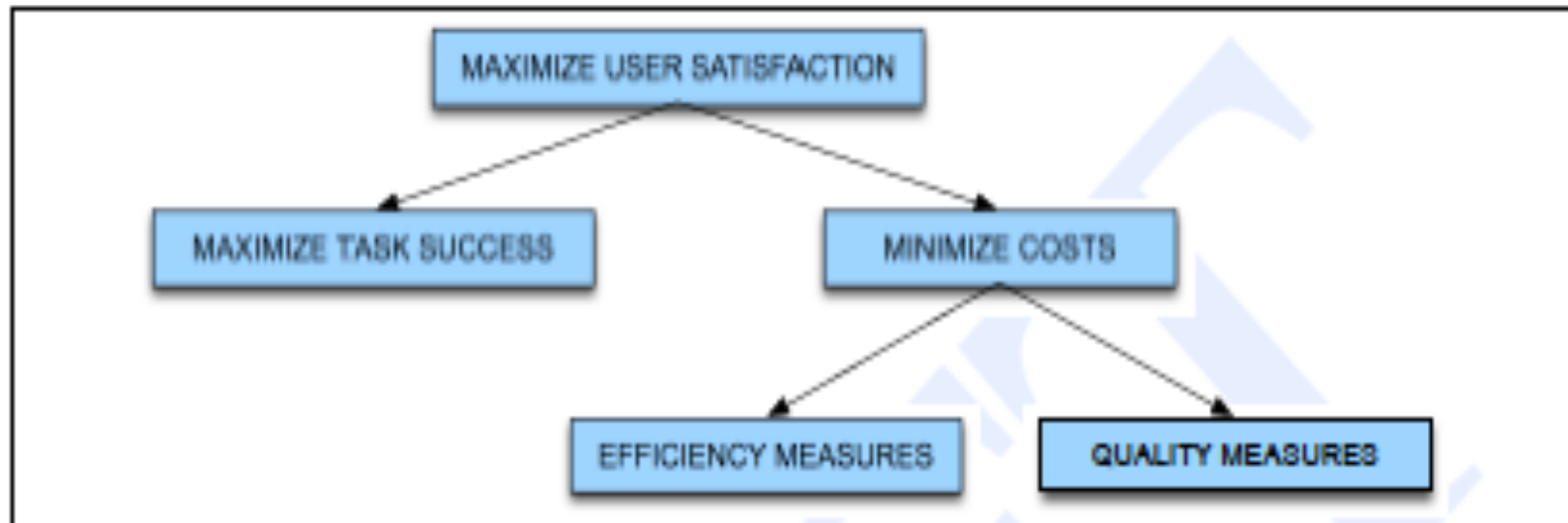


Figure 24.15 PARADISE's structure of objectives for spoken dialogue performance. After Walker et al. (1997).

PARADISE Model

- Compute user satisfaction with questionnaires
- 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

Now that we have a success metric

- Could we use it to help drive learning?
- In recent work we use this metric to help us learn an optimal **policy** or **strategy** for how the conversational agent should behave

A threshold is a human-designed policy!

- Could we learn what the right action is
 - Rejection
 - Explicit confirmation
 - Implicit confirmation
 - No confirmation
- By learning a policy which,
 - given various information about the current state,
 - dynamically chooses the action which maximizes dialogue success

Another strategy decision

- Open versus directive prompts
- When to do mixed initiative

- How we do this optimization?
- Markov Decision Processes

Summary

- Spoken Dialogue Systems:
 - Build on existing text-based NLP techniques, but
 - Incorporate dialogue specific factors:
 - Turn-taking, grounding, dialogue acts
 - Affected by computational and modal constraints
 - Recognition errors, processing speed, etc.
 - Speech transience, slowness
 - Becoming more widespread and more flexible