ulm university universität ulm

NACHRICHTENTECHNIK Dialogue Systems

Application and Evaluation of a Conditioned Hidden Markov Model for Estimating Interaction Quality of Spoken Dialogue Systems

Communications Technology – Dialogue Systems
Ulm University (Germany)
http://www.dialogue-systems.org

Stefan Ultes, Robert Elchab, Wolfgang Minker

Monday, December 03, 2012 – IWSDS 2012 – Ermenonville

Motivation

- Rapidly **growing importance** and use of Spoken Dialog Systems (SDS)
  - ✖ Still: static and error-prone

- Example:

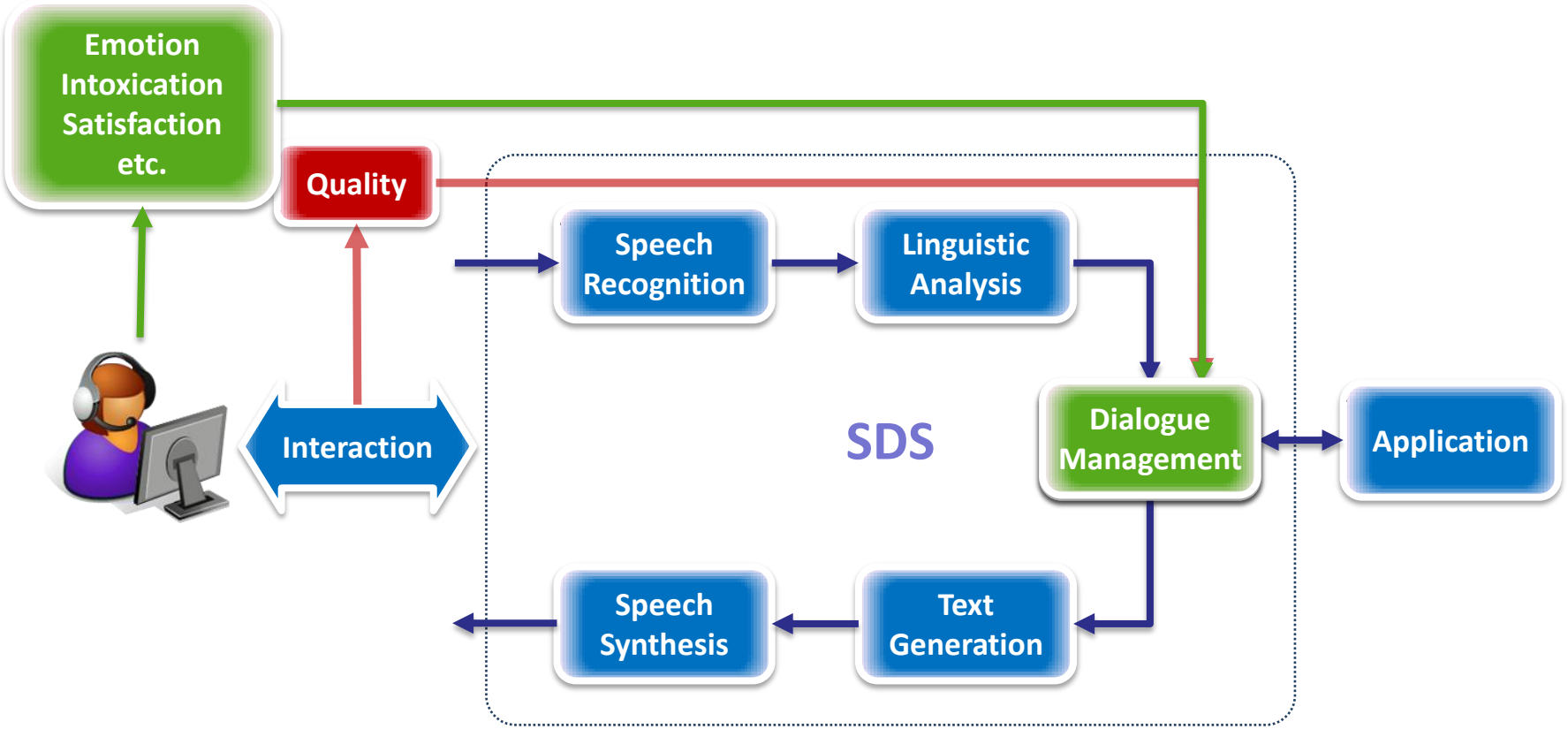
S: What can I do for you?
U: 54c Schedule
S: The 54C.
U: yes
S: Alright. Where do you want to leave from?
U: West North Avenue
S: Leaving from EIGHTH Avenue.
U: No , West North Avenue
S: Leaving from EIGHTH Avenue
U: No!

U: Arch Street and Cedar Avenue
S: [Leaving from NORTH Avenue.] Is this correct?
U: What?!
S: Leaving from FIFTH Avenue.
S: Where are you going?
U: I'm not leaving from FIFTH Avenue.
S: Going to EIGHTH Avenue. Is this correct?

<HANGUP>

Motivation

- Rapidly **growing importance** and use of Spoken Dialog Systems (SDS)
  - ✖ Still: static and error-prone



Motivation

- Requirements for Quality Metric:
  - Exchange-level quality measurement
    - Dialogue Management is performed after each system-user exchange
  - Automatically derivable
    - SDS are supposed to be autonomous

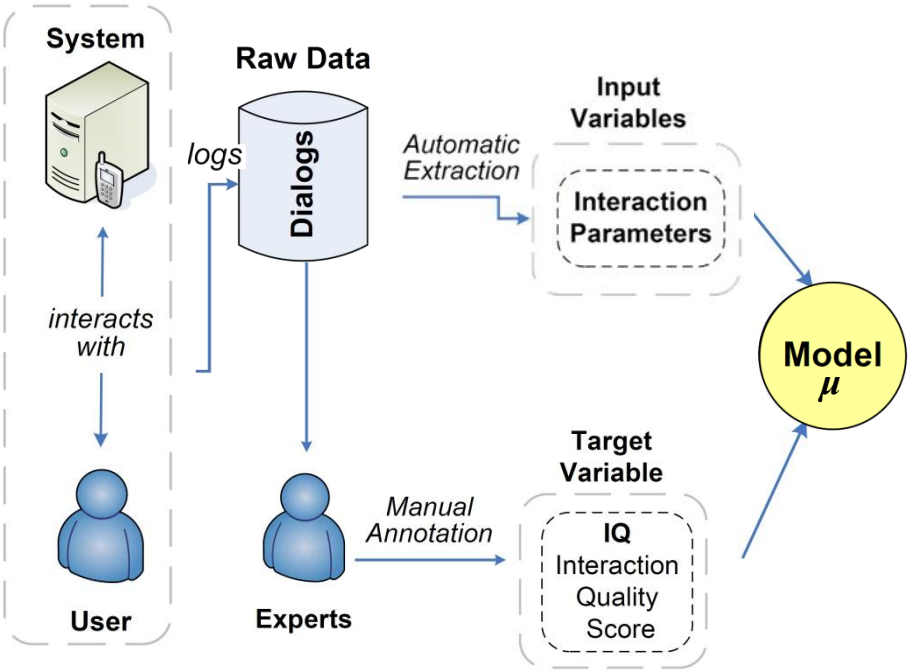
→ **Interaction Quality Paradigm** (Schmitt et al., SIGDial 2011)

Outline

- Motivation
- Interaction Quality Paradigm
- Conditioned Hidden Markov Model
- Experiment and Results
- Conclusion

Interaction Quality Paradigm

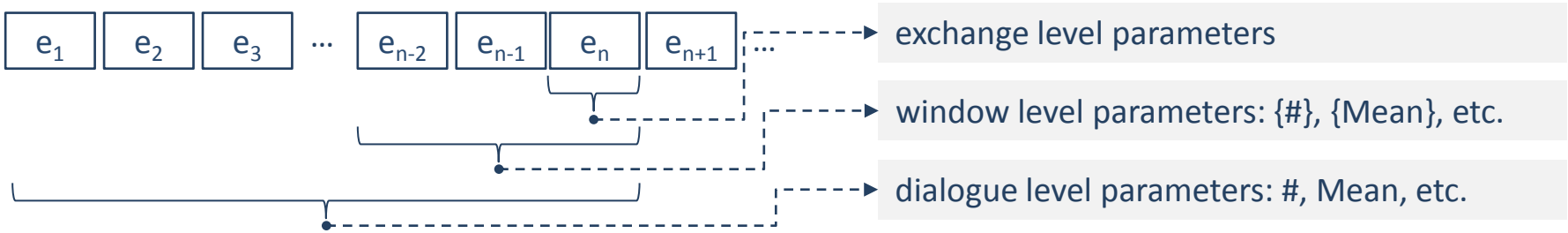
- Real system-user dialogues (Lets Go, CMU, Pittsburgh)
- Manual annotation of dialogues by three experts raters
  - Ratings from 5 (“satisfied”) to 1 (“extremely unsatisfied”)
- Automatic extraction of input variables
  - Generate statistical model to predict IQ on the exchange level



Interaction Quality Paradigm – Interaction Parameters

- Features on three levels:
  - Exchange level:

    Features derived from
    - Automatic Speech Recognition (ASR)
    - Spoken Language Understanding (SLU)
    - Dialogue Manager (DM) module
  - Window level:
    - Counts, means of exchange-level parameter
    - Computed over the last 3 exchanges
  - Dialogue level:
    - Counts, rates, means of exchange-level parameter
    - Computed over all exchanges up to the current turn

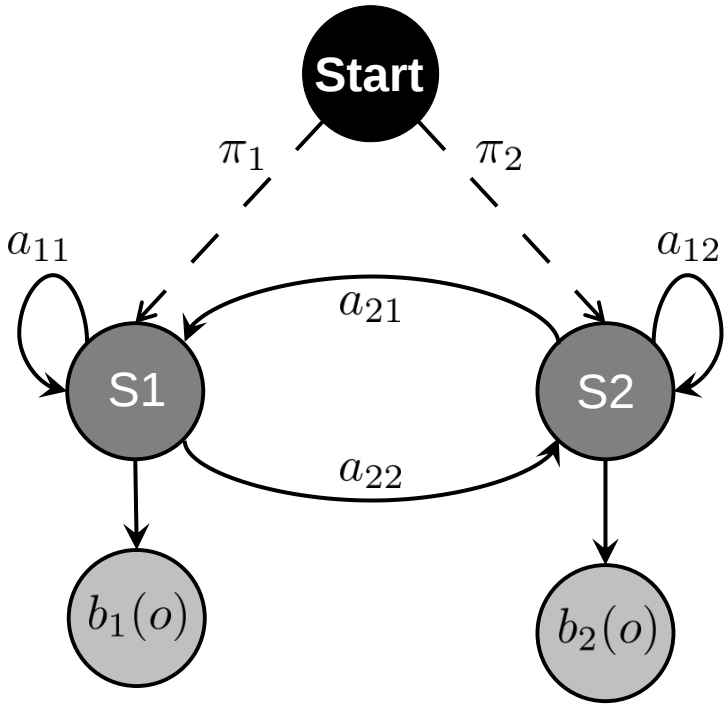


Statistical Models

- Schmitt et al.: Support Vector Machine
  - Problems:
    - Temporal dependencies only by design of special parameters
    - Previous IQ values not taken into account
- Our approach:
  - Models which take into account temporal dependencies and previous values inherently
    - Hidden Markov Model
    - Conditioned Hidden Markov Model

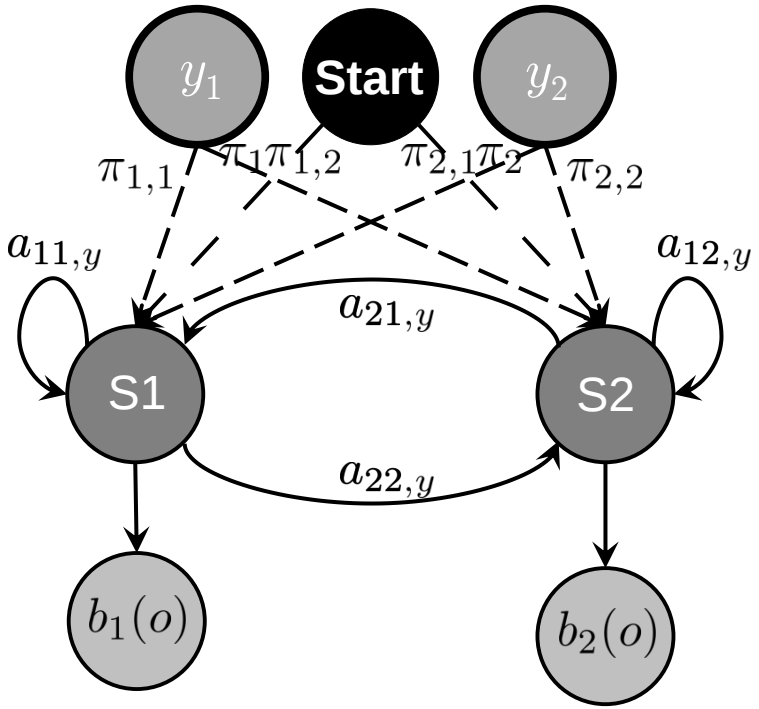
Hidden Markov Model (HMM)

- Often: one class per HMM
- Observation sequence probability $p(\vec{o}|\lambda)$



Conditioned Hidden Markov Model (CHMM) (Glodek et al., Interspeech 2011)

- Multiple classes per CHMM
- Class probability $p(y|\vec{o}, \lambda)$
- Training and evaluation algorithms must be altered



CHMM – Computing Class Probability

- Forward-Algorithm

$$\alpha_{t,y}(j) = b_j(\mathsf{o}^{(t)}) \cdot \sum_i a_{ij,y} \cdot \alpha_{t-1,y}(i)$$

$$\alpha_{1,y}(j) = b_j(\mathsf{o}^{(1)}) \cdot \pi_{j,y}$$

- Class probability

$$p(y|\mathsf{o}^{(n)}) = \frac{p(\mathsf{o}^{(n)}, y)}{\sum_y p(\mathsf{o}^{(n)}, y)} = \frac{p(\mathsf{o}^{(n)}|y) \cdot p(y)}{\sum_y (p(\mathsf{o}^{(n)}|y) \cdot p(y))}$$

$$p(\mathsf{o}^{(n)}|y) = \sum_i \alpha_{T,y}(i)$$

t – time step
i,j – hidden state number
y – label
b_j – observation probability
$a_{ij,y}$ – transition probability
$\pi_{j,y}$ – initial probability
$\mathsf{o}^{(n)}$ – observation sequence

NACHRICHTENTECHNIK Dialogue Systems

Experiment

- IQ-Recognition for Lets Go (LEGO corpus, Schmitt et al. 2012)
  - 200 calls
  - 4,885 system-user-exchanges
  - Each exchange manually annotated with IQ
- Evaluation metric: Unweighted Average Recall (UAR)
- Setup
  - 29 features per exchange (selected out of 46 available features)
  - 6-fold cross-validation
  - Classifiers **→ 9 states**
    - CHMM: Linear search for optimal number of hidden states
    - HMM: Five hidden states (one associated with one label)
    - SVM: Approach in accordance to Schmitt et al.
- Results

Classifier	CHMM	HMM	SVM
UAR	0.39	0.44	0.49

Conclusion

- SVM outperforms CHMM and HMM
  - Problem: Lack of data
    - CHMM
      - 5 classes, 9 states, 29 features → **8280 model parameters**
      - 3,908 training vectors (per fold)
    - HMM
      - 5 states, 29 features → **4380 model parameters**
      - 3,908 training vectors (per fold)
    - Less then 1 training vector per parameter
    - Can attribute selection help?

- Future work
  - Investigate influence of attribute selection
  - Investigate HMMs and CHMMs performance with more data
    - Data not available, has to be labeled

Thank you!