Cooperative User Models in Statistical Dialog Simulators

Meritxell González[1,2], Silvia Quarteroni[1], Giuseppe Riccardi[1] *and* Sebastian Varges[1]

[1] DISI - University of Trento - 38050 Povo, Italy

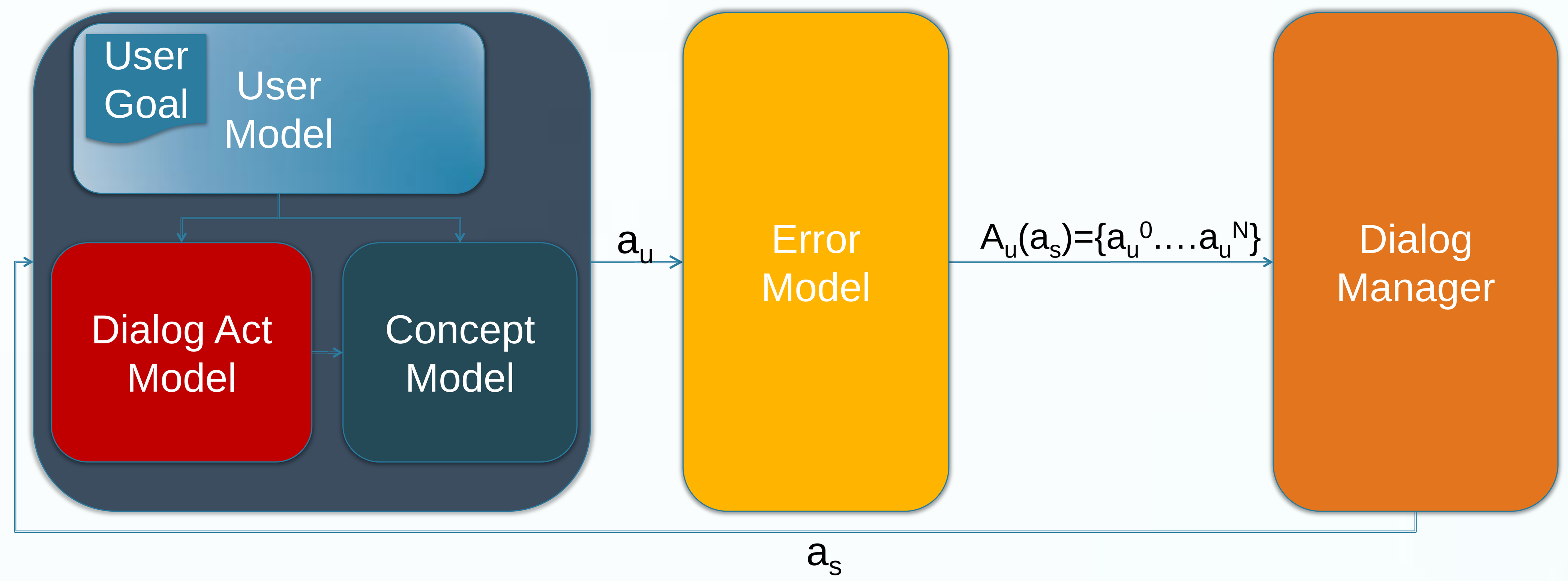
[2] TALP Center – Technical University of Catalonia – 08034 Barcelona, Spain

mgonzalez@lsi.upc.edu, {silviaq,riccardi,varges}@disi.unitn.it

Abstract

- Train a dialog simulator that combines traits of human behavior with domain-related aspects.
- A modular architecture for data-driven dialog simulation where the intentional component of user simulator includes a **User Model**.
- Combine the models for different features via the Expectation-Maximization algorithm.

Architecture



- Dialog Manager action: $a_s=\{da_0,\ldots,da_n\}$
- The User Simulator generates a N-best list of actions: $A_u(a_s)=\{a_u^0 \ldots a_u^N\}$
  - The *User Model* simulates a user behavior.
  - The *Dialog Act Model* generates a distribution of M actions. Each action has a probability based on the context as seen in the corpus.
  - The *Concept Model* generates concepts values according to the *User Goal.*
  - The *Error Model* distorts $\hat{a}_u$ to simulate the ASR-SLU noise.
- An action $\hat{a}_u$ is chosen following a random sampling according to the distributions as seen in the corpus.

User Model

- Transient features:
  - User Goal: task name and list of concepts and values
- Persistent features:
  - Patience = 1 - the tendency to abandon the conversation. *pat = 1 - P(HangUp | a_s)*
  - "noinput" probability accounts (also) for noisy environments. *noi = P(NoInput | a_s)*
  - Cooperativeness is the ratio of concepts mentioned in a_s that also appear in $\hat{a}_s$.

Cooperativeness Model and the context

- Cooperativeness feature:
  - *$coop_t$* is the rate of system and user DAs sharing the same concepts *within a turn*.

$$coop_t(a_u,a_s)=\frac{|a_u \cap a_s|}{|a_s|}$$

  - Feature *coop* is the dialog average of *$coop_t$.*

$$coop=\frac{\sum_{t=0}^{T} coop_t(a_u^i,a_s)}{\sum_{v=0}^{T}|a_s^v|}$$

- Cooperativeness model:
  1. Discretize *coop* into a binary variable: ***k*** **= high vs. low** cooperativeness.
     - *coop threshold in ADASearch is 0.28*
  2. Train a bigram model for high/low *coop.*

$A_u(a_s,k) = \{(a_u^0,P(a_u^0|a_s,k)),\ldots,(a_u^M,P(a_u^M|a_s,k))\}$

- Combining different models
  - **$A_u(a_s,k)$** is the cooperativeness model
  - **$A_u(a_s,\varphi)$** represents the context. It is the distribution of actions as obtained from the Dialog Act Model

$Au(a_s, \varphi) = \{(a_u^0,P(a_u^0|a_s, \varphi)),\ldots,(a_u^M,P(a_u^M|a_s, \varphi))\}$

- The cooperativeness and context model are linearly interpolated.

$A_u(a_s) = \lambda_k \cdot A_u(a_s,k) + \lambda_\varphi \cdot A_u(a_s, \varphi)$; $\lambda_\varphi + \lambda_k = 1$

- The weight of the cooperative model for a given bigram $[a_s,a_u^i]$ is calculated as follows:

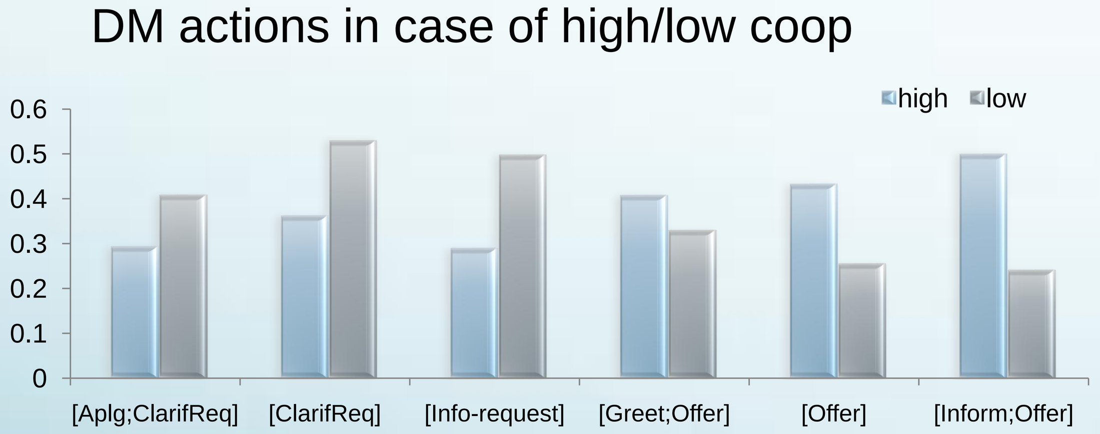
$$\lambda_k=\frac{\sum_{j=0}^{M} P(k|a_s,a_u^j)}{M}.$$

- And the global weight of the cooperative models is:

$$P(k|a_s,a_u^i)=\frac{P(a_u^i|a_s,k)}{P(a_u^i|a_s,k)+P(a_u^i|a_s,\varphi)}\ \forall_{i=0}^{M} a_u^i$$

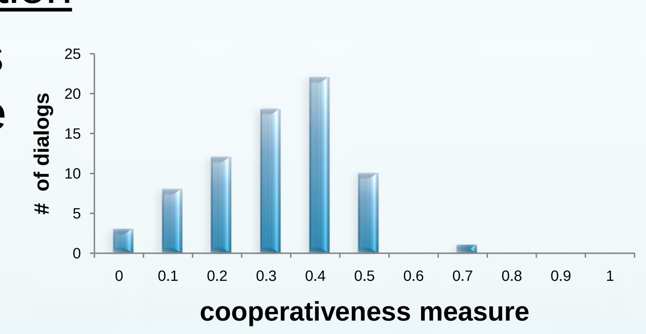
Experiments

- Effect of Cooperativeness
  - Estimated λ_k weights in response to selected DM actions in case of high/low coop



- λ_k achieves high values in case of uncooperative users in response to DM dialog acts as [ClarificationRequest] and [Info-request].
- In contrast, forward-looking actions, such as the ones including [Offer], seem to discard the contribution of the low coop model, but to favor the contribution provided by high coop.

- Cooperativeness distribution
  - The majority of dialogs is clustered around the [0.1..0.4] range.



- Dialog Act Precision & Recall wrt real calls
  - A simulated dialog act is correct when it appears in the real action a_u.

$$P_{DA}=\frac{|correct\ DA|}{|DA \in \hat{a}_u|}, R_{DA}=\frac{|correct\ DA|}{|DA \in \hat{a}_u|}$$

DA Model	Simulation (a^*_u)		Most Frequent (a^*_u)	
	P_{DA}	R_{DA}	P_{DA}	R_{DA}
OB	33.8	33.4	33.9	33.5
BI(+*coop*)	35.6 **(35.7)**	33.5 **(35.8)**	**49.3** (47.9)	**48.8** (47.4)
TP(+*coop*)	38.2 **(39.7)**	38.1 **(39.4)**	**51.1** (50.6)	**50.6** (50.2)

- Task Duration and Completion rates
  - We ran 60 simulated dialogs between the DM and each combination of the Task-based and Bigram models and high and low values of *coop*.
  - Under *high-coop* regime the number of turns taken to complete tasks is lower than under *low-coop*.
  - TCR is higher when cooperativeness is higher, indicating that *cooperative* users make dialogs not only shorter but also more efficient.

$$TCR=\frac{number\ of\ times\ a\ task\ has\ been\ completed}{total\ number\ of\ task\ requests}$$

	Lodging Enquiry		Lodging Reserv		Event Enquiry		All TCR
	#turns	TCR	#turns	TCR	#turns	TCR	TCR
OB	9.2±0.0	78%	9.7±1.4	82%	8.1±2.9	67%	77%
BI+low	15.1±4.1	71%	14.2±3.9	69%	9.3±1.8	52%	67%
BI+high	**12.1±2.5**	**75%**	**12.9±3.1**	**82%**	**7.8±1.8**	**75%**	**77%**
TB+low	13.6±4.1	76%	13.4±3.7	83%	8.4±3.3	65%	77%
TB+high	**11.6±2.8**	**80%**	**12.6±3.6**	**84%**	**6.5±1.9**	**57%**	**78%**
Real Dialogs	11.1±3.0	71%	12.7±4.7	70%	9.3±4.0	85%	73%

Conclusions

- Modular combination of user features with different models of dialog act, concept-value and ASR/SLU error simulation.
- Joint model of user intentions with a model of individual user traits.
- User's cooperativeness as a real-valued feature of the User Model .
- Improved accuracy of dialog act estimation when evaluated on a training set of real conversations.
- In future work, we will refine our simulators by studying more fine-grained and realistic User Model representations.

References

S. Quarteroni, M. González, G. Riccardi, S. Varges. Combining User Intention and Error Modeling for Statistical Dialog Simulators. INTERSPEECH 2010.

J. Schatzmann, K. Georgila, S. Young. Quantitative evaluation of user simulation techniques for spoken dialog systems. SIGDIAL 2005.

W. Eckert, E. Levin, R. Pieraccini. User modeling for spoken dialogue system evaluation. IEEE ASRU 1997.

F. Jelinek, R.L.Mercer. Interpolated estimation of Markov source parameters from sparse data. Pattern Recognition in Practice 1980.

S. Jung, C. Lee, K. Kim, G.G.LeeHybrid approach to user intention modeling for dialog simulation. ACL-IJCNLP 2009.

This work was partially funded by the European Commission - Marie Curie Excellence Grant for the ADAMACH project (contract No. 022593); and by the Spanish Ministry of Science and Innovation - OPENMT-2 project (TIN2009-14675-C03)