A Two-tier User Simulation Model for Reinforcement Learning of Adaptive Referring Expression Generation Policies

Srinivasan Janarthanam
School of Informatics
University of Edinburgh
s.janarthanam@ed.ac.uk

Oliver Lemon
School of Informatics
University of Edinburgh
olemon@inf.ed.ac.uk

Abstract

We present a new two-tier user simulation model for learning adaptive referring expression generation (REG) policies for spoken dialogue systems using reinforcement learning. Current user simulation models that are used for dialogue policy learning do not simulate users with different levels of domain expertise and are not responsive to referring expressions used by the system. The two-tier model displays these features, that are crucial to learning an adaptive REG policy. We also show that the two-tier model simulates real user behaviour more closely than other baseline models, using the `dialogue similarity` measure based on Kullback-Leibler divergence.

1 Introduction

We present a new user simulation model for learning adaptive referring expression generation (REG) policies for spoken dialogue systems using reinforcement learning methods. An adaptive REG policy equips a dialogue system to dynamically modify its utterances in order to adapt to user's domain knowledge level. For instance, to refer to domain objects, the system might use simple descriptive expressions with novices and technical jargon with experts. Such adaptations help grounding between the dialogue partners (Issacs and Clark, 1987). Since the user's knowledge level is unknown, the system must be able to adapt dynamically during the conversation. Hand-coding such a policy could be extremely difficult. (Janarthanam and Lemon, 2009b) have shown that such policies can be learned using simulation based reinforcement learning (RL) methods. The quality of such learned policies is directly dependent on the performance of the user simulations used to train them. So far, only hand-coded user simulations have been employed. In contrast, we now present a data driven two-tier user simulation model trained on dialogue data collected from real users. We also show that the two-tier model simulates real users more faithfully than other data driven baseline n-gram models (Eckert et al., 1997).

In section 2 we briefly discuss other work related to user simulations for dialogue policy learning using RL. In section 3 we describe the data used to build the simulation. Section 4 describes the simulation models in detail. In section 5 and 6 we present the evaluation metrics used and the results.

2 Related work

Several user simulation models have been proposed for dialogue management policy learning (Schatzmann et al., 2006; Schatzmann et al., 2007). However, these models cannot be directly used for REG policy learning because they interact with the dialogue system only using high-level dialogue acts. Also, they do not simulate different user groups like experts, novices, etc. In order to learn adaptive REG policies, user simulations need to respond to the system's choice of referring expressions and simulate user groups with different knowledge levels. We propose a two-tier simulation which simulates users with different knowledge levels and is sensitive to the system's choice of referring expressions.

Proceedings of SIGDIAL 2009: the 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue, pages 120–123,
Queen Mary University of London, September 2009. ©2009 Association for Computational Linguistics

3 Corpus

The "Wizard-of-Oz" (WOZ) methodology is a widely accepted way of collecting dialogue data for user simulation modeling (Whittaker et al., 2002). In this setup, real users interact with a human wizard disguised as a dialogue system. The wizard interprets the users responses and passes them on to the dialogue system. The dialogue system updates the dialogue state and decides the responses to user's moves. The task of the participant is to interact with the dialogue system to get instructions to setup a broadband Internet connection. The referring expression generation strategy is chosen before the dialogue starts and stays the same for the whole session. The strategies used were "jargon", "descriptive" and "tutorial". In the jargon strategy the system instructs the user using technical terms (e.g. "Plug the `broadband filter` into the `phone socket`."). In the descriptive strategy, it uses descriptive terms (e.g. "Plug the `small white box` into the `square white box on the wall`."). In the tutorial strategy, the system uses both jargon and descriptive terms together. The system provides clarifications on referring expressions when users request them. The participant's domain knowledge is also recorded during the task. Please refer to (Janarthanam and Lemon, 2009a) for a more details on our Wizard-of-Oz environment for data collection. The dialogues were collected from 17 participants (one dialogue each) with around 24 to 35 turns per dialogue depending on the strategy and user's domain knowledge.

4 User Simulation models

The dialogue data and knowledge profiles were used to build user simulation models. These models take as input the system's dialogue act $A_{s,t}$ (at turn t) and choice of referring expressions $REC_{s,t}$ and output the user's dialogue $A_{u,t}$ and environment $EA_{u,t}$ acts. User's observation and manipulation of the domain objects is represented by the environment act.

4.1 Advanced n-gram model

A simple approach to model real user behaviour is to model user responses (dialogue act and environment act) as advanced n-gram models (Georgila et al., 2006) based on many context variables - all referring expressions used in the utterance ($REC_{s,t}$), the user's knowledge of the REs (DK_u), history of clarification requests on the REs (H), and the system's dialogue act ($A_{s,t}$), as defined below:

$$P(A_{u,t}|A_{s,t}, REC_{s,t}, DK_u, H)$$

$$P(EA_{u,t}|A_{s,t}, REC_{s,t}, DK_u, H)$$

Although this is an ideal model of the real user data, it covers only a limited number of contexts owing to the limited size of the corpus. Therefore, it cannot be used for training as there may be a large number of unseen contexts which the model needs to respond to. For example, this model cannot respond when the system uses a mix of jargon and descriptive expressions in its utterance because such a context does not exist in our corpus.

4.2 A Two-tier model

Instead of using a complex context model, we divide the large context in to several sub-contexts and model the user's response based on them. We propose a two-tier model, in which the simulation of a user's response is divided into two steps. First, all the referring expressions used in the system's utterance are processed as below:

$$P(CR_{u,t}|RE_{s,t}, DK_{RE,u}, H_{RE}, A_{s,t})$$

This step is repeated for each expression $RE_{s,t}$ separately. The above model returns a clarification request based on the referring expression $RE_{s,t}$ used, the user's knowledge of the expression $DK_{RE,u}$, and previous clarification requests on the expression H_{RE} and the system dialogue act $A_{s,t}$. A clarification request is highly likely in case of the jargon strategy and less likely in other strategies. Also, if a clarification has already been issued, the user is less likely to issue another request for clarification. In such cases, the clarification request model simply returns `none`.

In the next step, the model returns a user dialogue act $A_{u,t}$ and an environment act $EA_{u,t}$ based on the system dialogue act $A_{s,t}$ and the clarification request $CR_{u,t}$, as follows:

$$P(A_{u,t}|A_{s,t}, CR_{u,t})$$

$$P(EA_{u,t}|A_{s,t}, CR_{u,t})$$

By dividing the complex context into smaller sub-contexts, the two-tier model simulates real users in contexts that are not directly observed in the dialogue data. The model will therefore respond to system utterances containing a mix of REG strategies (for example, one jargon and one descriptive expression in the same utterance).

4.3 Baseline Bigram model

A bigram model was built using the dialogue data by conditioning the user responses only on the system's dialogue act (Eckert et al., 1997).

$$P(A_{u,t}|A_{s,t})$$
$$P(EA_{u,t}|A_{s,t})$$

Since it ignores all the context variables except the system dialogue act, it can be used in contexts that are not observed in the dialogue data.

4.4 Trigram model

The trigram model is similar to the bigram model, but with the previous system dialogue act $A_{s,t-1}$ as an additional context variable.

$$P(A_{u,t}|A_{s,t}, A_{s,t-1})$$
$$P(EA_{u,t}|A_{s,t}, A_{s,t-1})$$

4.5 Equal Probability model baseline

The equal probability model is similar to the bigram model, except that it is not trained on the dialogue data. Instead, it assigns equal probability to all possible responses for the given system dialogue act.

4.6 Smoothing

We used Witten-Bell discounting to smooth all our models except the equal probability model, in order to account for unobserved but possible responses in dialogue contexts. Witten-Bell discounting extracts a small percentage of probability mass, i.e. number of distinct responses observed for the first time (T) in a context, out of the total number of instances (N), and redistributes this mass to unobserved responses in the given context ($V - T$) (where V is the number of all possible responses) . The discounted probabilities P^* of observed responses ($C(e_i) > 0$) and unobserved responses ($C(e_i) = 0$) are given below.

$$P^*(e_i) = \frac{C(e_i)}{N+T} \;\; if(C(e_i) > 0)$$
$$P^*(e_i) = \frac{t}{(N+T)(V-T)} \;\; if(C(e_i) = 0)$$

On analysis, we found that the Witten-Bell discounting assigns greater probability to unobserved responses than to observed responses, in cases where the number of responses per context is very low. For instance, in a particular context, the possible responses, their frequencies and their original probabilities were - `provide_info` (3, 0.75), `other` (1, 0.25), `request_clarification` (0, 0). After discounting, the revised probabilities P^* are 0.5, 0.167 and 0.33. `request_clarification` gets the whole share of extracted probability as it is the only unobserved response in the context and is more than the `other` responses actually observed in the data. This is counter-intuitive for our application. Therefore, we use a modified version of Witten-Bell discounting (given below) to smooth our models, where the extracted probability is equally divided amongst all possible responses. Using the modified version, the revised probabilities for the illustrated example are 0.61, 0.28 and 0.11 respectively.

$$P^*(e_i) = \frac{C(e_i)}{N+T} + \frac{T}{(N+T)V}$$

5 Metrics for evaluation of simulations

While there are many proposed measures to rank user simulation models with respect to real user data (Schatzmann et al., 2005; Georgila et al., 2006; Rieser and Lemon, 2006a; Williams, 2008), we use the `Dialogue Similarity` measure based on Kullback-Leibler (KL) (Cuayahuitl et al., 2005; Cuayahuitl, 2009) divergence to measure how similar the probability distributions of the simulation models are to the original real human data.

5.1 Dialogue Similarity

Dialogue Similarity is a measure of divergence between real and simulated dialogues and can measure how similar a model is to real data. The measure is based on Kullback-Leibler (KL) divergence and is defined as follows:

$$DS(P||Q) = \frac{1}{N}\sum_{i=1}^{N} \frac{D_{KL}(P||Q)+D_{KL}(Q||P)}{2}$$
$$D_{KL}(P||Q) = \sum_{i=1}^{M} p_i * log(\frac{p_i}{q_i})$$

The metric measures the divergence between distributions P and Q in N different contexts with M responses per context. Ideally, the dialogue similarity between two similar distributions is close to zero.

6 Evaluation results

We consider the Advanced N-gram model to be a realistic model of the real human dialogue corpus, as it takes into account all context variables and is reasonably smoothed to account for unobserved user responses. Therefore, we compare the probability distributions of all the other models to

Model	$A_{u,t}$	$EA_{u,t}$
Two-tier	0.078	0.018
Bigram	0.150	0.139
Trigram	0.145	0.158
Equal Probability	0.445	0.047

Table 1: Dialogue Similarity with Modified Witten-Bell discounting w.r.t Advanced N-gram model

the advanced n-gram model using the `dialogue similarity` measure. The results of the evaluation are given in table 1.

The results show that the two-tier model is much closer (0.078, 0.018) to the Advanced N-gram model than the other models. This is due to the fact that the bigram and trigram models don't take into account factors like the user's knowledge, the strategy used, and the dialogue history. By effectively dividing the RE processing and the environment interaction, the two-tier simulation model is not only realistic in observed contexts but also usable in unobserved contexts (unlike the Advanced N-gram model).

7 Conclusion

We have presented a data driven user simulation model called the two-tier model for learning REG policies using reinforcement learning. We have also shown that the two-tier model is much closer to real user data than the other baseline models. We will now train REG policies using the two-tier model and test them on real users in the future.

Acknowledgements

The research leading to these results has received funding from the EPSRC (project no. EP/E019501/1) and from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 216594 (CLASSiC project `www.classic-project.org`), and from the British Council's UKERI programme.

References

H. Cuayahuitl, S. Renals, O. Lemon, and H. Shimodaira. 2005. Human-Computer Dialogue Simulation Using Hidden Markov Models. In *Proc. of ASRU 2005*.

H. Cuayahuitl. 2009. *Hierarchical Reinforcement Learning for Spoken Dialogue Systems*. Ph.D. thesis, University of Edinburgh, UK.

W. Eckert, E. Levin, and R. Pieraccini. 1997. User Modeling for Spoken Dialogue System Evaluation. In *Proc. of ASRU97*.

K. Georgila, J. Henderson, and O. Lemon. 2006. User Simulation for Spoken Dialogue System: Learning and Evaluation. In *Proc of ICSLP 2006*.

E. A. Issacs and H. H. Clark. 1987. References in conversations between experts and novices. *Journal of Experimental Psychology: General*, 116:26–37.

S. Janarthanam and O. Lemon. 2009a. A Wizard-of-Oz environment to study Referring Expression Generation in a Situated Spoken Dialogue Task. In *Proc. ENLG'09*.

S. Janarthanam and O. Lemon. 2009b. Learning Lexical Alignment Policies for Generating Referring Expressions for Spoken Dialogue Systems. In *Proc. ENLG'09*.

V. Rieser and O. Lemon. 2006a. Cluster-based User Simulations for Learning Dialogue Strategies. In *Proc. Interspeech/ICSLP*.

J. Schatzmann, K. Georgila, and S. J. Young. 2005. Quantitative Evaluation of User Simulation Techniques for Spoken Dialogue Systems. In *Proc. SIGdial workshop on Discourse and Dialogue '05*.

J. Schatzmann, K. Weilhammer, M. N. Stuttle, and S. J. Young. 2006. A Survey of Statistical User Simulation Techniques for Reinforcement Learning of Dialogue Management Strategies. *Knowledge Engineering Review*, pages 97–126.

J. Schatzmann, B. Thomson, K. Weilhammer, H. Ye, and S. J. Young. 2007. Agenda-based User Simulation for Bootstrapping a POMDP Dialogue System. In *Proc of HLT/NAACL 2007*.

S. Whittaker, M. Walker, and J. Moore. 2002. Fish or Fowl: A Wizard of Oz Evaluation of Dialogue Strategies in the Restaurant Domain. In *Language Resources and Evaluation Conference*.

J. Williams. 2008. Evaluating User Simulations with the Cramer-von Mises Divergence. *Speech Communication*, 50:829–846.