

# Learning Dialogue Strategies from Older and Younger Simulated Users

**Kallirroï Georgila**  
Institute for Creative Technologies  
University of Southern California  
Playa Vista, USA  
kgeorgila@ict.usc.edu

**Maria K. Wolters**  
School of Informatics  
University of Edinburgh  
Edinburgh, UK  
maria.wolters@ed.ac.uk

**Johanna D. Moore**  
School of Informatics  
University of Edinburgh  
Edinburgh, UK  
J.Moore@ed.ac.uk

## Abstract

Older adults are a challenging user group because their behaviour can be highly variable. To the best of our knowledge, this is the first study where dialogue strategies are learned and evaluated with both simulated younger users and simulated older users. The simulated users were derived from a corpus of interactions with a strict system-initiative spoken dialogue system (SDS). Learning from simulated younger users leads to a policy which is close to one of the dialogue strategies of the underlying SDS, while the simulated older users allow us to learn more flexible dialogue strategies that accommodate mixed initiative. We conclude that simulated users are a useful technique for modelling the behaviour of new user groups.

## 1 Introduction

State-of-the-art statistical approaches to dialogue management (Frampton and Lemon, 2006; Williams and Young, 2007) rely on having adequate training data. Dialogue strategies are typically inferred from data using Reinforcement Learning (RL), which requires on the order of thousands of dialogues to achieve good performance. Therefore, it is no longer feasible to rely on data collected with real users. Instead, training data is generated through interactions of the system with simulated users (SUs) (Georgila et al., 2006). In order to learn good policies, the behaviour of the SUs needs to cover the range of variation seen in real users (Georgila et al., 2006; Schatzmann et al., 2006). Furthermore, SUs are critical for evaluating candidate dialogue policies.

To date, SUs have been used to learn dialogue strategies for specific domains such as flight reser-

vation, restaurant recommendation, etc., and to learn both how to collect information from the user (Frampton and Lemon, 2006) as well as how to present information to the user (Rieser and Lemon, 2009; Janarthanam and Lemon, 2009). In addition to covering different domains, SUs should also be able to model relevant user attributes (Schatzmann et al., 2006), such as cooperativeness vs. non-cooperativeness (López-Cózar et al., 2006; Jung et al., 2009), or age (Georgila et al., 2008). In this paper, we focus on user age.

As the proportion of older people in the population increases, it becomes essential to make spoken dialogue systems (SDS) easy to use for this group of people. Only very few spoken dialogue systems have been developed for older people (e.g. Nursebot (Roy et al., 2000)), and we are aware of no work on learning specific dialogue policies for older people using SUs and RL.

Older people present special challenges for dialogue systems. While cognitive and perceptual abilities generally decline with age, the spread of ability in older people is far larger than in any other segment of the population (Rabbitt and Anderson, 2005). Older users may also use different strategies for interacting with SDS. In our previous work on studying the interactions between older and younger users and a simulated appointment scheduling SDS (Wolters et al., 2009b), we found that some older users were very “social”, treating the system like a human, and failing to adapt to the SDS’s system-initiative dialogue strategy. A third of the older users, however, tended to be more “factual”, using short commands and conforming to the system’s dialogue strategy. In that, they were very similar to the younger users (Wolters et al., 2009b).

In previous work (Georgila et al., 2008), we successfully built SUs for both older and younger

adults from the corpus used by (Wolters et al., 2009b) and documented in (Georgila et al., 2010). When we evaluated the SUs using metrics such as precision and recall (Georgila et al., 2006; Schatzmann et al., 2006), we found that SUs trained on older users’ data can cover behaviour patterns typical of younger users, but not the opposite. The behaviour of older people is too diverse to be captured by a SU trained on younger users’ data. This result agrees with the findings of (Wolters et al., 2009b; Georgila et al., 2010).

In this study, we take our work one step further—we use the SUs developed in (Georgila et al., 2008) to learn dialogue policies and evaluate the resulting policies with data from both older and younger users. Our work is important for two reasons. First, to the best of our knowledge this is the first time that people have used SUs and RL to learn dialogue strategies for the increasingly important population of older users. Second, despite the fact that SUs are used for learning dialogue strategies it is not clear whether they can learn policies that are appropriate for different user populations. We show that SUs can be successfully used to learn policies for older users that are adapted to their specific patterns of behaviour, even though these patterns are far more varied than the behaviour patterns of younger users. This provides evidence for the validity of the user simulation methodology for learning and evaluating dialogue strategies for different user populations.

The structure of the paper is as follows: In section 2 we describe our data set, discuss the differences between older and younger users as seen in our corpus, and describe our user simulations. In section 3, we present the results of our experiments. Finally, in section 4 we present our conclusions and propose future work.

## 2 The Corpus

In the original dialogue corpus, people were asked to schedule health care appointments with 9 different simulated SDS in a Wizard-of-Oz setting. The systems varied in the number of options presented at each stage of the dialogue (1, 2, 4), and in the confirmation strategies used (explicit confirmation, implicit confirmation, no confirmation). System utterances were generated using a simple template-based algorithm and synthesised using a female Scottish English unit selection voice. The human Wizard took over the function of speech recognition (ASR), language understanding (NLU), and dialogue management com-

ponents. No ASR or NLU errors were simulated, because having to deal with ASR and/or NLU errors in addition to task completion would have increased cognitive load (Wolters et al., 2009a).

The system (Wizard) followed a strict policy which resulted in dialogues with a fixed schema: First, users arranged to see a specific health care professional, then they arranged a specific half-day, and finally, a specific half-hour time slot on that half-day was agreed. Users were not allowed to skip any stage of the dialogue. This design ensured that all users were presented with the relevant number of options and the relevant confirmation strategy at least three times per dialogue. In a final step, the Wizard confirmed the appointment.

The full corpus consists of 447 dialogues; 3 dialogues were not recorded. A total of 50 participants were recruited, of which 26 were older, aged between 50 and 85 years, and 24 were younger, aged between 18 and 30 years. The older users contributed 232 dialogues, the younger ones 215. Older and younger users were matched for level of education and gender. All dialogues were transcribed orthographically and annotated with dialogue acts and dialogue context information. Using a unique mapping, we associate each dialogue act with a  $\langle$ speech act, task $\rangle$  pair, where the speech act is task independent and the task corresponds to the slot in focus (health professional, half-day or time slot). For example,  $\langle$ confirm\_pos, hp $\rangle$  corresponds to positive explicit confirmation of the health professional slot. For each dialogue, detailed measures of dialogue quality were recorded: objective task completion, perceived task completion, appointment recall, length (in turns), and extensive user satisfaction ratings. For a detailed discussion of the corpus, see (Georgila et al., 2010).

The choice of dialogue strategy did not affect task completion and appointment recall, but had significant effects on efficiency (Wolters et al., 2009a). Task completion and appointment recall were the same for older and younger users, but older users took more turns to complete the task (Wolters et al., 2009a). Clear differences between the two user groups emerge when we look at interaction patterns in more detail (Wolters et al., 2009b; Georgila et al., 2010). Older people tend to “ground” information (using repetitions) and take the initiative more than younger people. In our corpus it was very common that the older person would provide information about the half-day and the time slot of the appointment before having been asked by the system. However, due to the

	Experiment 1	Experiment 2
slot filled	+50	+50
appointment confirmed	+200	+200
dialogue length	-5 per turn	-5 per turn
slot confirmed	+100	not used
wrong order	-500	not used

Table 1: Reward functions for the experiments.

strict policy of the Wizard, this information would be ignored and the system would later ask for the information that had already been provided.

In our SUs, each user utterance corresponds to a user action described by a list of ⟨speech act, task⟩ pairs. There are 31 distinct system actions and 389 distinct actions for older users. Younger people used a subset of 125 of the older users’ actions. Our SUs do not simulate ASR or NLU errors since such errors were not simulated in the collection of the corpus.

We built  $n$ -grams of system and user actions with  $n$  varying from 2 to 5. Given a history of  $n-1$  actions from system and user, the SU generates an action based on a probability distribution learned from the training data (Georgila et al., 2006). In the present study,  $n$  was set to 3, which means that each user action is predicted based on the previous user action and the previous system action.

### 3 Learning Dialogue Strategies

We performed two experiments. In Experiment 1, our goal was to learn the policy of the Wizard, i.e. the strict system-initiative policy of requesting and confirming information for each slot before moving to the next slot, in the following order: health professional, half-day, time slot. In Experiment 2, our goal was to learn a more flexible policy that could accommodate some degree of user initiative.

The reward functions for both experiments are specified in Table 1; they are similar to the reward functions used in the literature, e.g. (Frampton and Lemon, 2006). Slots that have been filled successfully and confirmed appointments are rewarded, while long dialogues are penalised. For Experiment 1, policies were rewarded that filled slots in the correct order and that confirmed each slot after it had been filled. A large penalty was imposed when the policy deviated from the strict slot order (health professional, half-day, time slot). For Experiment 2, these constraints were removed. Slots could be filled in any order. Confirmations were not required because there was no speech act in the corpus for confirming more than one slot at a time.

In both experiments we used the SARSA- $\lambda$  algorithm (Sutton and Barto, 1998) for RL. 30,000 iterations were used for learning the final policy for each condition. For each experiment, we learned two policies, Policy-Old, which was based on simulated older users, and Policy-Young, which was based on simulated younger users. The resulting policies were then tested on simulated older users (Test-Old) and simulated younger users (Test-Young). To have comparable results between Experiment 1 and Experiment 2, during testing we score our policies using the reward function of Experiment 2. The best possible score is 190, i.e. the user fills all the slots in one turn and then confirms the appointment. (Note that +50 points are given when a slot is only filled, not confirmed too.) For each test condition, we generated 10,000 simulated dialogues. Overall scores for each combination of policy and SU were established using 5-fold cross-validation.

Our results are summarised in Figure 1. While average rewards were not affected by policy type (ANOVA,  $F(1, 68)=1$ ,  $p=0.3$ ) or training data set ( $F(1, 185)=3$ ,  $p=0.09$ ), we found a very strong interaction between policy type and data set ( $F(1, 3098)=51$ ,  $p=0.000$ ). Learning with simulated younger users yields better strict policies than learning with older users (Tukey’s Honest Significant Difference Test,  $\Delta=20$ , 95% CI = [11, 30],  $p=0.000$ ), while learning with simulated older users yields better flexible policies than learning with younger users ( $\Delta=15$ , 95% CI = [6, 24],  $p=0.001$ ). This is what we would expect from our corpus analysis, since the interaction behaviour of older users is far more variable than that of younger users (Wolters et al., 2009b; Georgila et al., 2010).

The strict policy that was learned from simulated younger users was as follows, with only slight variations: first request the type of health professional, then implicitly confirm the health professional and request the half-day slot, then implicitly confirm the half-day slot and request the time slot, and then confirm the appointment. The strict policy learned from simulated older users was similar, but less successful, because most older users do not readily conform to the fixed structure.

The flexible policy learned from simulated older users takes into account initiative from the user and does not always confirm. The score for the flexible policy learned from simulated younger users was relatively low, even though the resulting

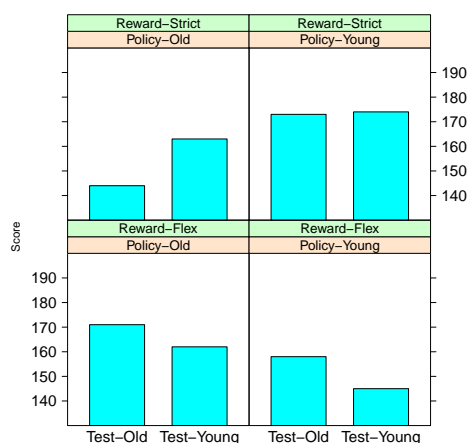


Figure 1: Mean scores for each combination of reward function, training set, and test set (5-fold cross-validation).

policy was very similar to the strict policy learned from younger users (i.e. a sequence of information requests and implicit confirmations), and even though the behaviour of younger users is far more predictable than the behaviour of older users. It appears that the explicit penalty for violating the order of slots is crucial for fully exploiting the patterns in younger users' behaviour.

## 4 Conclusions

We have shown that SUs can be used to learn appropriate policies for older adults, even though their interaction behaviour is more complex and diverse than that of younger adults. Crucially, simulated older users allowed us to learn a more flexible version of the strict system-initiative dialogue strategies that were used for creating the original corpus of interactions. These results are consistent with previous analyses of the original corpus (Wolters et al., 2009b; Georgila et al., 2010) and support the validity of the user simulation methodology for learning and evaluating dialogue strategies.

In our future work, we will experiment with more complex SUs, e.g. linear feature combination models (Georgila et al., 2006), and see if they can be used to learn similar policies. We also plan to study the effect of training and testing with different user simulation techniques, such as  $n$ -grams versus linear feature combination models.

## Acknowledgements

This research was partially supported by the MATCH project (SHEFC-HR04016, <http://www.match-project.org.uk>).

Georgila is supported by the U.S. Army Research, Development, and Engineering Command (RDECOM). The content does not necessarily reflect the position or the policy of the U.S. Government, and no official endorsement should be inferred.

## References

- M. Frampton and O. Lemon. 2006. Learning more effective dialogue strategies using limited dialogue move features. In *Proc. ACL*.
- K. Georgila, J. Henderson, and O. Lemon. 2006. User simulation for spoken dialogue systems: Learning and evaluation. In *Proc. Interspeech*.
- K. Georgila, M. Wolters, and J. Moore. 2008. Simulating the behaviour of older versus younger users. In *Proc. ACL*.
- K. Georgila, Maria Wolters, J.D. Moore, and R.H. Logie. 2010. The MATCH corpus: A corpus of older and younger users' interactions with spoken dialogue systems. *Language Resources and Evaluation*, 44(3):221–261.
- S. Janarathanam and O. Lemon. 2009. A two-tier user simulation model for reinforcement learning of adaptive referring expression generation policies. In *Proc. SIGdial*.
- S. Jung, C. Lee, K. Kim, and G.G. Lee. 2009. Hybrid approach to user intention modeling for dialog simulation. In *Proc. ACL*.
- R. López-Cózar, Z. Callejas, and M. McTear. 2006. Testing the performance of spoken dialogue systems by means of an artificially simulated user. *Artificial Intelligence Review*, 26(4):291–323.
- P. Rabbitt and M.M. Anderson. 2005. The lacunae of loss? Aging and the differentiation of human abilities. In F.I. Craik and E. Bialystok, editors, *Lifespan Cognition: Mechanisms of Change*, chapter 23. Oxford University Press, New York, NY.
- V. Rieser and O. Lemon. 2009. Natural language generation as planning under uncertainty for spoken dialogue systems. In *Proc. EACL*.
- N. Roy, J. Pineau, and S. Thrun. 2000. Spoken dialog management for robots. In *Proc. ACL*.
- J. Schatzmann, K. Weilhammer, M. Stuttle, and S. Young. 2006. A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies. *Knowledge Engineering Review*, 21(2):97–126.
- R.S. Sutton and A.G. Barto. 1998. *Reinforcement Learning: An Introduction*. MIT Press.
- J. Williams and S. Young. 2007. Partially observable Markov decision processes for spoken dialog systems. *Computer Speech and Language*, 21(2):393–422.
- M. Wolters, K. Georgila, J.D. Moore, R.H. Logie, S.E. MacPherson, and M. Watson. 2009a. Reducing working memory load in spoken dialogue systems. *Interacting with Computers*, 21(4):276–287.
- M. Wolters, K. Georgila, J.D. Moore, and S.E. MacPherson. 2009b. Being old doesn't mean acting old: How older users interact with spoken dialog systems. *ACM Trans. Accessible Computing*, 2(1).