Towards Learning Transferable Conversational Skills using Multi-dimensional Dialogue Modelling

Simon Keizer and Verena Rieser
Interaction Lab
Heriot-Watt University, Edinburgh (UK)
{s.keizer,v.t.rieser}@hw.ac.uk

Abstract

Statistical approaches to dialogue management have brought improvements in robustness and scalability of spoken dialogue systems, but still rely heavily on in-domain data, thus limiting their cross-domain scalability. In this paper, we present a new multi-dimensional, statistical dialogue management framework, in which transferable conversational skills can be learnt by separating out domain-independent dimensions of communication. Our preliminary experiments demonstrate the effectiveness of such transfer.

1 Introduction

Virtual personal assistants, such as Siri, Cortana, Google Now, and Alexa, have made commercial use of interactive spoken language technology. However, commercial exploitation of advanced spoken dialogue technology requires new methods for cost-effective development and efficient adaptation to new domains. We argue that this problem can be tackled by taking a *multi-dimensional* approach, which is based on the idea that in addition to an underlying task/activity, dialogue participants simultaneously address several other aspects of communication when interpreting and generating utterances, such as giving and eliciting feedback, following social conventions, and managing turn-taking and timing. In the example below, the user both greets the system and asks for a cheap Indian restaurant, before releasing the turn; the system then takes the turn and indicates that it needs some time to retrieve the requested information; in the second part the system both provides this information and gives feedback about understanding the user's question (underlined).

| Usr: | *Hello, I am looking for a cheap <u>Indian</u> restaurant* |
| | SOCIAL:GREET; TASK:INFORM; TURN:RELEASE |

| Sys: | *Let me see, …* |
| | TURN:TAKE; TIME:PAUSING; TASK:INFORMSEARCH |

| Sys: | *The Rice Boat is an <u>Indian</u> restaurant in the <u>cheap</u> pricerange* |
| | AUTO-FEEDBACK:INFORM; TASK:INFORM |

Following this notion of multi-dimensionality of dialogue as described by Bunt (2011) and early exploratory work on multi-dimensional dialogue management by Keizer and Bunt (2006, 2007), we present a new framework for statistical dialogue management which explicitly accounts for these different dimensions of communication. By separating out domain-independent dimensions, our approach has the potential to learn a set of transferable conversational skills, enabling more efficient cross-domain adaptation.

2 Multi-dimensional dialogue manager

Following general design features of the POMDP systems described in (Young et al., 2010) and (Thomson and Young, 2010), we created a generic dialogue management framework, consisting of state monitoring and action selection components, and an agenda-based user simulator and error model for testing, training and evaluation. In contrast to existing POMDP-based systems, dialogue contributions here are modelled in terms of dialogue acts from the ISO 24617-2 multi-dimensional dialogue act taxonomy (ISO, 2012), and the action selection component consists of multiple dialogue act agents, each dedicated to generating candidate dialogue acts from one dimension. The agents are modelled as MDPs and can be trained simultaneously using (multi-agent) reinforcement learning (currently Monte Carlo control with linear value function approxi-

mation). With our new multi-dimensional framework, we can train a multi-agent dialogue manager in a particular domain, resulting in a domain specific policy and several domain-independent policies, which can be re-used and adapted in a new domain with the aim to speed up learning.

3 Preliminary experiments in simulation

As a first proof-of-concept experiment, we have developed a multi-dimensional dialogue manager for the restaurant information domain, consisting of three dialogue act agents, corresponding to the dimensions *Task* (5 actions, including asking for user preferences, making recommendations, presenting restaurant information), *AutoFeedback* (3 actions, including asking clarification questions), and *SocialOblMan* (2 actions, including goodbye acts). In all policy optimisation experiments, 10 independent training runs have been carried out, and the evaluation results are averages over the 10 corresponding policy evaluations. All policies were trained over 40k dialogues with an exploration rate linearly decaying from $\epsilon = 0.4$ to $\epsilon = 0$ and a fixed learning rate of $\alpha = 0.001$. The agents shared a single reward function (+30 upon task completion; -1 per turn).

Each of the three learning curves in Fig. 1 shows the performance of trained policies at different training stages, where each data point represents the average reward over 3000 evaluation dialogues (averaged over 10 policies). The red curve with square markers corresponds to the baseline system described above that was trained from scratch.
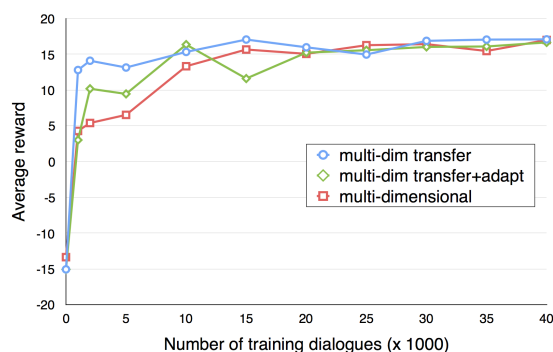

Figure 1: Policy evaluation results in terms of average success rate at different training stages.

After jointly optimising the three MDP policies, two domain-independent policies have been obtained that have the potential to be re-used in a new domain. To demonstrate this potential in a first preliminary test, we re-trained the dialogue man-

ager by retaining the trained *AutoFeedback* and *SocOblMan* policies (as if they were trained in a different source domain) and training only the task policy from scratch (for the 'new' target domain). This domain transfer exercise was carried out in two settings: 1) *multi-dim transfer*: only updating the task policy, i.e., keeping the trained domain-independent policies fixed, and 2) *multi-dim transfer+adapt*: updating all three polices during training, i.e., adapting the trained domain-independent policies to the 'new' domain. The effectiveness of domain transfer is demonstrated by the corresponding learning curves in Fig. 1, which show improved performance levels at the earlier stages of training in comparison to the non-transferred multi-dimensional system. Setting 1 (blue, with circular markers) shows clear and consistent improvement, whereas the improvement in setting 2 (green, with diamond markers) is more modest and training seems less stable.

4 Conclusion and Future Work

We have presented the first implementation of a multi-dimensional statistical dialogue manager and illustrated our approach with proof-of-concept experiments in simulation, demonstrating the feasibility of training transferable conversational skills using multi-agent reinforcement learning. We will extend our dialogue manager to support a wider range of dialogue act combinations, and are building an end-to-end system for the restaurant and smart home domains, in order to demonstrate our results on real data and across domains.

References

H. Bunt. 2011. Multifunctionality in dialogue. *Computer Speech & Language*, 25(2):222–245.

ISO. 2012. *ISO 24617-2 Language resource management – Semantic annotation framework – Part 2: Dialogue acts*. International Organization for Standardization, Geneva, Switzerland.

S. Keizer and H. Bunt. 2007. Evaluating combinations of dialogue acts for generation. In *Proc. SIGdial*.

S. Young, M. Gašić, S. Keizer, F. Mairesse, J. Schatzmann, B. Thomson, and K. Yu. 2010. The Hidden Information State Model: a practical framework for POMDP-based spoken dialogue management. *Computer Speech & Language*, 24(2):150–174.