Learning Culture-Specific Dialogue Models from Non Culture-Specific Data

Kallirroi Georgila and David Traum

Institute for Creative Technologies, University of Southern California,
12015 Waterfront Drive, Playa Vista, CA 90094, United States
{kgeorgila, traum}@ict.usc.edu

Abstract. We build culture-specific dialogue policies of virtual humans for negotiation and in particular for argumentation and persuasion. In order to do that we use a corpus of non-culture specific dialogues and we build simulated users (SUs), i.e. models that simulate the behavior of real users. Then using these SUs and Reinforcement Learning (RL) we learn negotiation dialogue policies. Furthermore, we use research findings about specific cultures in order to tweak both the SUs and the reward functions used in RL towards a particular culture. We evaluate the learned policies in a simulation setting. Our results are consistent with our SU manipulations and RL reward functions.

Keywords: dialogue systems, culture-specific dialogue models, reinforcement learning, simulated users, negotiation, argumentation, persuasion.

1 Introduction

Virtual humans are artificial agents that have a humanlike appearance and behavior. Virtual humans often engage in conversations and can play a number of roles, for example, negotiate with humans or other virtual humans. Having virtual humans with explicit models of specific cultures can be very effective for teaching culture-specific skills because it creates a realistic setting for interaction. Also, having culture-specific virtual humans can make it easier for people of that culture to interact with and understand the virtual humans [12].

In this paper we focus on building culture-specific dialogue policies of virtual humans in negotiation and in particular in argumentation and persuasion, i.e. dialogue policies that dictate what kind of arguments and persuasion strategies the virtual human will use to accomplish its goal depending on the cultural behavior that we want to simulate. This task is particularly challenging for two reasons:

1. There is a shortage of culture-specific dialogue data in negotiation but also in other domains. What we know about culture-specific behavior is usually the result of surveys in which people from different cultures are asked to give their opinions on several matters.
2. Although these surveys can provide valuable culture-specific information it is not clear how their findings can translate into culture-specific models of conversational behavior.

Note that in this paper when we refer to culture-specific dialogue models we do not mean models of specific real cultures (e.g. Americans versus Chinese) but of dimensions on which cultures are known to vary. Brett and Gelfand [1] identified three aspects in cross-cultural negotiation: individualism versus collectivism, egalitarianism versus hierarchy, and low context versus high context communication. Typically Western individuals are individualistic, egalitarian, and use low context communication while Eastern individuals are collectivistic, hierarchical, and use high context communication.[1] In this paper we focus on individualism and altruism in particular, but the ideas and techniques can be applied to other types of cultural dimensions, such as collectivism.

In order to learn dialogue models of cultural dimensions from data not specific to these dimensions we propose the following novel approach. We use a corpus of dialogues not specific to any dimension and we build simulated users (SUs), i.e. models that simulate the behavior of real users [4, 7, 8]. Then using these SUs and Reinforcement Learning (RL) [3, 5, 6, 11, 13] we learn negotiation policies. Furthermore, we tweak both the SUs and the reward functions used in RL towards a particular cultural dimension, by taking into account research findings about the cultural dimensions of interest. Our research contribution is two-fold:

1. To date, statistical approaches to dialogue management based on RL have focused on information slot-filling applications (e.g. tourist information domains) [4, 13], largely ignoring other types of dialogue with rare exceptions [3, 6]. Here, we use RL for learning negotiation (argumentation and persuasion) policies. As we will see in the following, this is a particularly challenging task due to the complexity of the dialogue state and the large number of system and user actions.
2. With our approach we can learn dialogue policies for a specific cultural dimension without having dialogue data specific to that dimension. Furthermore, unlike Heeman [6] who built hand-crafted SUs we learn our initial SUs from a corpus and then tweak them. While the idea of manipulating the SUs to simulate different types of users is not new [7, 8], to our knowledge the idea of manipulating the reward functions towards a particular behavior is rather novel. In recent work Georgila et al. [5] manipulated the reward functions to learn strict versus flexible system policies for appointment booking but that approach was limited in the sense that it did not involve using different sets of actions for each policy that we want to learn as we do here.

The structure of the paper is as follows: In section 2 we briefly introduce the concepts of RL and SUs. In section 3 we present the corpus used in our experiments. In section 4 we describe how we build our SUs from our corpus and how we tweak them towards a particular cultural dimension. In section 5 we present how we use the SUs built in section 4 in order to learn culture-specific negotiation policies. In section 6, we describe our evaluation experiments. Then in section 7 we discuss our findings together with ideas for future work, and finally in section 8 we present our conclusions.

[1] In high-context cultures the listener must understand the contextual cues in order to grasp the full meaning of the message. In low-context cultures communication tends to be specific, explicit, and analytical.

2 Reinforcement Learning and Simulated Users

In the Reinforcement Learning (RL) paradigm, managing a dialogue can be seen as a Markov Decision Process (MDP) or a Partially Observable Markov Decision Process (POMDP) where dialogue moves transition between dialogue states and rewards are given at the end of a successful dialogue. The solution to the dialogue management problem is a policy specifying for each state the optimal action to take. Typically rewards depend on the domain and can include factors such as task completion, dialogue length, and user satisfaction.

Several research groups have investigated the use of RL for dialogue management in slot-filling dialogues, including [4, 5, 13]. Slot-filling dialogues are dialogues in which the user presents a complex query or service request (e.g. a hotel booking), and the system iteratively asks for more information to fully specify and confirm a set of "slots" that are needed to generate a database query (e.g. location, price range, room type) and ultimately satisfy the user's request. Dialogue policy decisions are typically whether to ask for a slot value, confirm a slot value, query the database, or present an answer. A typical reward function is to multiply the number of slots that have been filled and confirmed by a weighting factor (e.g. 100 points) and subtract the number of system turns multiplied by a weighting factor (e.g. 5 points) [5].

In contrast to slot-filling dialogue, in negotiation dialogue the system and the user have opinions about the optimal outcomes and try to reach a joint decision. Dialogue policy decisions are typically whether to present, accept, or reject a proposal, whether to compromise, etc. Rewards may depend on the type of policy that we want to learn. For example, a cooperative policy should be rewarded for accepting the other party's proposals. On the other hand a non-cooperative policy should be rewarded for ignoring the other party's proposals. Unlike slot-filling dialogues, the use of RL for learning negotiation dialogue policies has only recently been investigated [3, 6]. More specifically, Heeman [6] reported work on representing the RL state for learning negotiation dialogue policies for a furniture layout task.

The problem with RL is that it 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) [4]. In order to learn good policies, the behavior of the SUs needs to cover the range of variation seen in real users [4]. Furthermore, SUs are critical for evaluating candidate dialogue policies [8].

3 Our Corpus

In our negotiation domain, the data consists of dialogues between American undergraduates playing the role of a florist and a grocer who share a retail space. The dialogues were collected by Laurie R. Weingart, Jeanne M. Brett, and Mary C. Kern at Northwestern University. The florist and the grocer negotiate on four issues: the design of the space, the temperature, the rent, and their advertising policy. The florist and the grocer have different goals, preferences, and use different types of arguments.

We have annotated 21 dialogues using a cross-cultural argumentation and persuasion annotation scheme that we have developed.

This scheme is an adaptation of existing coding schemes on negotiation [2, 9, 10], following a review of literature on cross-cultural differences in negotiation styles (e.g. [1, 14]), and our observations from its application to coding negotiation dialogues in different domains. To our knowledge this is the first annotation scheme designed specifically for coding cross-cultural argumentation and persuasion strategies. Previous work on cross-cultural negotiation [1] has not focused on argumentation or persuasion in particular.

Table 1 depicts an example dialogue annotated with our coding scheme. Actually the annotations are more complex but here they are simplified for brevity since their presentation is outside the scope of the paper.

As mentioned above, in the corpus the florist and the grocer negotiate about four issues and sometimes these issues can be intertwined. For example, there could be trade-offs such as "I will agree on design A if you agree on a low temperature". Given that the task of learning dialogue policies with RL can be very complex even for simple slot-filling applications, in this initial experiment we decided to simplify the problem as much as possible. Thus we focus on learning how to negotiate about only one of the issues, the temperature. The florist is in favor of lower temperatures to keep her flowers fresh whereas the grocer prefers higher temperatures so that her customers feel comfortable.

Table 1. Example annotated dialogue with speech acts in the florist-grocer domain.

Speaker	Utterance	Speech Act
Florist	How does that work for you?	request_info.preference
Grocer	Well, personally for the grocery I think it is better to have a high temperature.	provide_argument.logic
Grocer	Just because I want the customers to feel comfortable.	elaborate
Florist	Okay.	acknowledge
Grocer	And also if it is warm, people are more apt to buy cold drinks to keep themselves comfortable and cool.	elaborate
Florist	That's true.	accept
Florist	But what about your products staying fresh? Don't they have to stay fresh or otherwise?	rebut_argument.logic

So we created a new smaller corpus by extracting the parts related to the temperature issue from the original corpus. We also excluded all dialogues with intertwined issues (3 dialogues) and dialogues where one party makes an offer in the first turn and the other party agrees immediately (3 dialogues). Thus we ended up with 15 shorter dialogues. Furthermore, we simplified the speech acts as shown in Table 2. These simplified dialogues were used for training our SUs as we will see in section 4. Also, Table 3 shows some statistics of the simplified corpus used in our experiment.

Table 2. Example simplified dialogue used for training the SUs.

Simplified Speech Acts
florist, provide_info.preference
florist, release_turn
grocer, provide_argument
grocer, offer
grocer, release_turn
florist, reject
florist, release_turn
grocer, provide_argument
grocer, elaborate
grocer, offer
grocer, release_turn
florist, accept
florist, release_turn

Table 3. Statistics of the simplified corpus used for training the SUs.

	Florist	Grocer	Total
Total # turns	65	65	130
Avg # turns per dialogue	4.3	4.3	8.7
Total # utterances	87	101	188
Avg # utterances per dialogue	5.8	6.7	12.5

4 Simulated Users

Our SUs are built on the speech act level from dialogues in the format depicted in Table 2. Note that we have inserted one more action "release_turn", which was not part of the original corpus to mark the boundaries between turns. Our SUs are based on n-grams of speech acts [4]. For example, valid 3-grams (Table 2) would be:

- grocer,provide_argument grocer, elaborate → grocer,offer
- florist,provide_info.preference florist,release_turn → grocer,provide_argument

The first 3-gram indicates that if the grocer provides an argument and then elaborates on this argument, then a possible action is for the grocer to make an offer. The second 3-gram indicates that if the florist provides her preference on the temperature and then releases the turn, then a possible action is for the grocer to provide an argument. The probability of each action is computed from our corpus. In this experiment we used 3-grams. The list of SU actions (as well as system actions) is given in Table 4. As we can see, our annotated dialogue data does not include information about cultural dimensions such as individualism. Thus we cannot directly learn from the corpus a SU of a particular cultural dimension. In our experiment we consider two different types of SUs, an individualist SU that never compromises, and an altruist SU that is the exact opposite of an individualist. The individualist SU-florist always generates arguments in favor of low temperatures, offers low

temperatures, rejects high temperatures, and so forth. The altruist SU-florist always generates arguments in favor of high temperatures, offers high temperatures, rejects low temperatures, and so forth. Likewise for the individualist and altruist SU-grocers.

Table 4. System policy and SU actions used in our experiment.

System and SU Actions
request_info.preference
provide_info.preference
provide_argument
elaborate
rebut_argument
acknowledge
offer
accept
reject
release_turn

5 Learning Negotiation Policies

After we have built our SUs, we have these SUs interact with our system (i.e. a virtual human) using RL in order to learn different policies. A virtual human that learns by interacting with a SU tweaked to care about individual gain, is expected to learn how to negotiate better against this type of conversational interlocutor. To ensure that our virtual human will also learn to simulate a particular cultural dimension we manipulate the reward functions used in RL. For example, a virtual human that cares about individual gain will always be rewarded for actions that lead to individual gain and penalized for actions that lead to individual loss or mutual gain. More specifically we consider two types of policies in the same fashion as for the SUs. Thus the individualistic florist policy is rewarded when the outcome of the conversation is agreement on a low temperature (+800 points) and penalized otherwise (-800 points). The altruistic florist policy is rewarded when the outcome of the negotiation is agreement on a high temperature (+800 points) and penalized otherwise (-800 points). Likewise for the individualistic and altruistic grocer policies. To facilitate learning we have also added one more penalty (-800 points) for some incoherent sequences of actions, i.e. when the action “elaborate” or “rebut_argument” appears before a “provide_argument” and when an “accept” or “reject” action appears when no offers or arguments are on the table. There is also a penalty of -10 points for each policy and SU action. The fastest possible successful dialogue can be for one of the interlocutors to make an offer and the other to accept. Thus the highest possible reward in a dialogue can be 800 minus 4 actions = 760, the four actions are “offer”, “release_turn”, “accept”, “release_turn”. Table 5 shows the reward functions used in our experiment. The goal of RL is to learn the optimal action in each dialogue state so that the desired outcome is achieved (e.g. a high temperature for the individualist grocer, a high temperature for the altruist florist, etc.).

Another issue is how to represent the state so that the problem is tractable and at the same time good policies can be learned. In this paper we used the state representation shown in Table 6, which leads to 864 possible states. We can see each feature with all the possible values it can take. Finally the policy actions are the same as the SU actions (see Table 4).

Table 5. Reward functions for each type of policy.

Type of Policy	Outcome	Incoherent Sequence	Penalty per Action
Individualist florist	low +800	-800	-10
Individualist florist	high -800	-800	-10
Altruist florist	low -800	-800	-10
Altruist florist	high +800	-800	-10
Individualist grocer	low -800	-800	-10
Individualist grocer	high +800	-800	-10
Altruist grocer	low +800	-800	-10
Altruist grocer	high -800	-800	-10

Table 6. State representation for learning.

State Representation
Current speaker (florist/grocer)
Most recent temperature supported by the florist (low/high)
Most recent temperature supported by the grocer (low/high)
Is there an argument on the table and by whom? (none/florist/grocer)
Is there an offer on the table and by whom? (none/florist/grocer)
If there is an offer, what is the temperature offered? (low/high)
Is there a rejected offer (the most recent rejection) and by whom? (none/florist/grocer)
If there is a rejected offer, what is the rejected temperature? (low/high)

For training we used the SARSA-λ algorithm [11] with greedy exploration at 30% to explore the state-action pair space. We ran 20,000 iterations for learning the final policy for each condition. More specifically, we learned an individualistic florist policy trained against both an individualist SU-grocer and an altruist SU-grocer, an altruistic florist policy trained against both an individualist SU-grocer and an altruist SU-grocer, and so forth. All possible combinations are shown in Table 7 in the evaluation section (section 6).

6 Evaluation of Learned Negotiation Policies

We evaluate our learned policies against our SUs. In some cases these are the SUs used for training which can be a potential problem, and certainly is an issue to be addressed in future work (for example when a florist policy is trained with an individualist SU-grocer and also tested with an individualist SU-grocer). However, due to data sparsity we cannot perform cross-validation. The data would not be enough for training reliable SUs.

We run each policy against all types of SUs (2000 simulated dialogues) and we report the outcome (how many successes we have, how many failures, and how many dialogues end with no agreement). For the individualistic florist policy the dialogue is considered successful if the final agreed temperature is low, in other words if the result of the negotiation favors the florist, and so forth. Results are given in Table 7. The notation is as follows: FI(GA)-GI stands for an individualistic florist policy trained against an altruist SU-grocer and tested against an individualist SU-grocer, and so forth.

The policies perform well (either they win or there is a tie) when interacting (in testing) with the SUs of the opposite culture (e.g. individualistic policy versus altruist SU), which is a good result. When this is not the case either there is no agreement or the SU wins. It is not surprising that the policy does not win in those cases but that the SU wins instead of having a tie (as in interactions FA(GA)-GA and GI(FI)-FI). This is an issue for further investigation. Another interesting issue is that one would expect that a policy trained on a SU of the same cultural dimension, e.g. FI(GI), would not perform well because in training there would always be disagreements. But these policies sometimes behave well, i.e. FI(GI)-GA, FA(GA)-GI, GI(FI)-FA, and GA(FA)-FI, but obviously not as well as FI(GA)-GA, FA(GI)-GI, GI(FA)-FA, and GA(FI)-FI respectively.

Table 7. Evaluation results for all combinations of policies and SUs.

Type of Policy	# Successes	# No Agreements	# Failures
FI(GI)-GI	0	536	1464
FI(GI)-GA	1772	228	0
FI(GA)-GI	0	547	1453
FI(GA)-GA	1804	196	0
FA(GI)-GI	1792	208	0
FA(GI)-GA	0	534	1466
FA(GA)-GI	1454	546	0
FA(GA)-GA	0	2000	0
GI(FI)-FI	0	2000	0
GI(FI)-FA	1332	668	0
GI(FA)-FI	0	685	1315
GI(FA)-FA	1661	339	0
GA(FI)-FI	1701	299	0
GA(FI)-FA	0	706	1294
GA(FA)-FI	1287	713	0
GA(FA)-FA	0	1375	625

7 Discussion

Our results are generally consistent with our SU probability manipulations and reward functions, which is encouraging. However, in order to make the problem tractable and learn these policies we had to compromise in many respects. The question that arises is what kind of improvements could be done while at the same time keeping the learning task tractable.

In this experiment in order to keep things tractable we make the assumption that there is no middle-ground behavior, which is unrealistic. In a real setting, it would make sense for the agents to compromise in some cases especially when the individualist florist interacts with the individualist grocer or the altruist florist interacts with the altruist grocer.

Furthermore, the SU-florist or SU-grocer always support one temperature each. In the future we intend to allow the SUs to generate arguments about different temperatures based on a probability distribution and the dialogue context. For example, the individualist SU-grocer will generate arguments in favor of high temperatures with a much higher probability than arguments in favor of middle-ground temperatures. That will lead to more realistic simulations because in our data there are cases where the florist or the grocer provide arguments in favor of their interlocutor or of a middle-ground solution.

We have also limited the number of actions to learn only to 10 and have kept the dialogue state small for tractability (for example we do not take into account the previous actions, which is a very important feature). All these compromises of course affect the quality of the learned policies. In future work we will investigate different state representations and action sets and see how they affect performance. We will also evaluate the policies against one another, not only against SUs.

Finally, the metric that we use for our evaluations is rather crude and it does not give us any insight about what happens in the course of the dialogue (the same is true for metrics that measure the success of RL-based policies as the number of slots that are filled and confirmed in slot-filling applications [4, 5, 13]), but we believe that it is a good first step towards developing evaluation metrics for new types of dialogue other than slot-filling dialogues.

8 Conclusions

We built culture-specific dialogue policies of virtual humans in negotiation and in particular in argumentation and persuasion. In order to do that we used a corpus of non-culture specific dialogues and built SUs. Then using these SUs and RL we learned negotiation dialogue policies. Furthermore, we took into account research findings about specific cultures in order to tweak both the SUs and the reward functions used in RL towards a particular culture. We evaluated the learned policies in a simulation setting. Our results are consistent with our SU manipulations and RL reward functions.

Acknowledgments

This research was funded by a MURI award through ARO grant number W911NF-08-1-0301. We are grateful to Laurie R. Weingart, Jeanne M. Brett, and Mary C. Kern who provided us with the florist-grocer dialogues. We also thank Ron Artstein, Angela Nazarian, Michael Rushforth, and Katia Sycara for their contribution to the development of the coding manual, which was used for annotating our corpus. The corpus was annotated by Angela Nazarian.

References

1. Brett, J.M., Gelfand, M.J.: A cultural analysis of the underlying assumptions of negotiation theory. In: L. Thomson (ed) Frontiers of Negotiation Research, Psychology Press, pp. 173--201 (2006)
2. Carnevale, P.J., Pruitt, D.G., Seilheimer, S.D.: Looking and competing: Accountability and visual access in integrative bargaining. Journal of Personality and Social Psychology, 40(1), 111--120 (1981)
3. English, M.S., Heeman, P.A.: Learning mixed initiative dialogue strategies by using reinforcement learning on both conversants. In: Proc. of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP), Vancouver, Canada, pp. 1011--1018 (2005)
4. Georgila, K., Henderson, J., Lemon, O.: User simulation for spoken dialogue systems: Learning and evaluation. In: Proc. of the International Conference on Spoken Dialogue Processing (Interspeech-ICSLP), Pittsburgh, PA, USA, pp. 1065-1068 (2006)
5. Georgila, K., Wolters, M.K., Moore, J.D.: Leaning dialogue strategies from older and younger simulated users. In: Proc. of the Annual SIGdial Meeting on Discourse and Dialogue (SIGdial), Tokyo, Japan, pp. 103--106 (2010)
6. Heeman, P.A.: Representing the reinforcement learning state in a negotiation dialogue. In: Proc. of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Merano, Italy, pp. 450--455 (2009)
7. Jung, S., Lee, C., Kim, K., Lee, G.G.: Hybrid approach to user intention modeling for dialog simulation. In: Proc. of the Annual Meeting of the Association for Computational Linguistics (ACL), Suntec, Singapore, pp. 17--20 (2009)
8. López-Cózar, R., Callejas, Z., McTear, M.: Testing the performance of spoken dialogue systems by means of an artificially simulated user. Artificial Intelligence Review, 26(4), 291--323 (2006)
9. Pruitt, D.G., Lewis, S.A.: Development of integrative solutions in bilateral negotiation. Journal of Personality and Social Psychology, 31(4), 621--633 (1975)
10. Sidner, C.L.: An artificial discourse language for collaborative negotiation. In: Proc. of the National Conference on Artificial Intelligence, Cambridge, MA, USA, pp. 814--819 (1994)
11. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. MIT Press (1998)
12. Traum, D.: Cultural models for virtual humans. In: Proc. of Human Computer Interaction International (HCII) (2009)
13. Williams, J.D., Young, S.: Scaling POMDPs for spoken dialogue management. IEEE Trans. on Audio, Speech and Language Processing, 15(7), 2116--2129 (2007)
14. Zaharna, R.S.: Understanding cultural preferences of Arab communication partners. Public Relations Review, 21(3), 241--255 (1995)