Reviewer 6 of IROS 2012 submission 358

Comments to the author
======================

Firstly, I generally like the approach to address human-robot-interaction problems using learning, and particularly the author's stated goal of allowing diverse inputs. I also understand that this is a difficult problem, and that simplifications are necessary to make initial steps. However, in this case, I can definitely say that I believe the simplification chosen in this paper has gone too far. I will further outline this in the following.

Secondly, I consider the particular problem addressed in this paper as generally solved. This does not mean that further work is of no interest, but that it must be compared properly to the existing work. Specifically, I would like to point the author's to the work of Heckman et al., "Teaching a Humanoid Robot: Headset-Free Speech Interaction for Audio-Visual Association Learning", published at RO-MAN 2009 and Bolder et.al, "Expectation-driven Autonomous Learning and Interaction System", published at HUMANOIDS 2008. These two papers present a very similar learning system which notably includes verb learning.

Now, coming back to my belief that the system presented has been simplified too far. I must say that for a person working in interactive systems, the combination of the statements "an interactive learning system with a flexible protocol" (section I, end) and "We start by assuming that the teacher provides a simple binary feedback" (II.A), is quite funny, really. How is a binary feedback flexible? Quite to the contrary, it is usually considered the most restrictive form of feedback possible!

I understand that this is only meant to be the start, but the only other example given (learning of action labels) is not much better.

Regarding terminology: A "protocol" describes the possible relations of how signals can occur. What you learn is not the protocol, you learn the signals.

In particular, the setup described completely ignores how the user should know what the robot can expect. If the user must know beforehand when feedback is understood, and that only binary feedback or labels are possible, the value of being able to choose the exact wording of such feedback, while not useless, is greatly diminished. For a discussion of what people usually say in similar situations, you might want to consider Peltason et al. "Talking with robots about object: A system-level evaluation in HRI", HRI 2012, which explicitly set out to accommodate diverse behavior. Their system is not learned, but may still be insightful to see what a more realistic interaction should contain. btw, I'd like to note that the problem of words being presented in isolation is shared with the work by Heckman et al mentioned above. This is illustrative: The real problem for learning in interaction is how to fit the various parts of an interaction /together/. Learning the pieces is comparatively (!) trivial.
Regarding this part of the paper's claims, I must conclude that the goal of a "flexible protocol" has not been achieved, and that the claims of the authors in this regard must be removed.
 Moreover, I also consider the word learning setup to be oversimplifying. In particular, by considering only isolated words, it ignores the problem of co-articulation: The modification of phonetic expression by the previous and succeeding words in a sentence, which causes much more difficult variation than isolated repetition. Therefore, demonstrating that 20 isolated words can be learned from the available input does not in any way demonstrate that the same could be achieved when the words are presented in the context of a sentence.

Furthermore, in a realistic setting, it is not at all clear which of the words in a sentence is most relevant for the given learning task. There are means to guess (e.g., emphasis, compare Schillingmann et al., ICDL 2011), but this introduces much more confusion than a simple binary confusion (as considered by the authors in section III.B 3).

Therefore, I must conclude that the goal of learning arbitrary feedback signals is only partially demonstrated.

Now, considering the action learning, I commend the authors for combining all the various pieces into a working system. That said, the reported value of more than a 100 repetitions is /much/ more than can be tolerated by a human interaction partner. In this vein, I would like to point the authors to more recent work by Cakmak et. al. than they have considered so far, which explicitly addresses this problem, firstly, "Designing Robot Learners that Ask Good Questions" by Cakmak and Thomaz, HRI 2012, and secondly, "Trajectories and Keyframes for Kinesthetic Teaching: A Human-Robot Interaction Perspective.", by Akgun et al, HRI 2012. Crucially, in their work, the robot becomes active and can ask questions based on an Active Learning approach. A similar mixed-initiative approach (without learning) has also been presented in Lütkebohle et al "The Curious Robot: Structuring interactive object learning", ICRA 2009. This doesn't have to be considered for the present paper, but for future work, I would strongly encourage the authors to consider an active component for their system. It's really essential to bring down the number of examples required, and also helps users during interaction.

As a last note: The paper is not bad, but most of the strong statements made abouts its flexibility are not appropriate, and really had me annoyed. You may know learning very well, but regarding interaction, I'd say you still have quite a bit to learn. Keep up the good work, but tone down the claims.
##
##
Reviewer 5 of IROS 2012 submission 358

Comments to the author
======================

This paper describes an approach to jointly learning the meaning of feedback terms (signifying positive or negative reward) as well as learning a task using the learned reward signal. The approach is based on expectation maximization. Overall, the idea is an interesting one, but I found the presentation confusing.

First, I'm surprised that the EM algorithm in equations 2-4 only depends on the current state and action. Doesn't it make more sense to iterate back through the previous trials as well? Second, Algorithm 2 seems to be the key result in the paper, and it is built around building up a distribution over tasks, q_k. However, I am not clear on how this distribution is represented or updated; this distribution is not mentioned anywhere in the text. Finally, there are very many grammatical errors that made the paper difficult to understand.

It would be useful if the paper gave an intuitive explanation of why the algorithm works. What regularities about the feedback signal is it exploiting in order to learn word meanings as well as the task metric? What are the limitations? Where will it start to break down. Also, it would be interesting if they reported the performance at recognizing the feedback signal; presumably it's going up over time, or well the system would not learn the task, but it would be useful to actually see it happening.

Reviewer 1 of Humanoids 2012 submission 173

Comments to the author
======================

The main contribution of the paper is a human-robot interaction, where the robot learns the task based on unknown teaching signals. The robot learns simultaneously what the task is and what the meaning of the unknown teaching signals is. These methods are well developed and generally clearly explained. Simulation results are provided showing different properties of the proposed system. The proposed algorithm was implemented on real robot as well.

The writing of the paper is in general good.

There are several instances of important details that are unexplained:
1. How does the robot know where the cube is? Did you use came for perception? What would happen if human (teacher) moves the cube to another location?
2. P.4-5. Task Representation: It is not clear how the rewards are given. Are they given to the current state, i.e. where the robot is at the moment, or only when the robot reaches the target state?
3. It is not clear when the teacher should use the prior knowledge, i.e. red and green button. Are there any conditions for that? Or it is completely random?
4. It is not clear to me why the robot needs more iteration if it is guided by teacher compare to the feedback example? One would expect the opposite. Can you explain this in more details?
5. In fig. 8 authors show that if buttons + signals are used, the system learns faster. Why is that, do you have any explanation? Aren't there more possible answers if you use both of them?

Comments on the Video Attachment
================================

The video in general looks nice. However I suggest you shorten it.

Reviewer 2 of Humanoids 2012 submission 173

Comments to the author
======================

This paper presents a robot learning algorithm which learns a task and meanings of teaching signals like voice feedbacks or guidance commands. Therefore, the learning procedure is designed in a incremental learning manner. It is an interesting work to show learning a task under noisy and unknown teaching/feedback signals under social human robot interaction setting. However, some details are missing. In order to improve the manuscript, the reviewer would like to suggest the followings.

It is questionable how to extend the proposed approach to more complex scenarios from a very simple setup shown in Sec. III experimental result. In this paper the world is presented as a discrete state set. How to deal with continuous world? How to relax many assumptions used in the paper (knowing that possible feedbacks are either correct or wrong)?

Please indicate the complexity, the computational cost.

Although the general algorithm is addressed as an EM algorithm, the details are often unclear. What are the exact model parameters (theta)? Please clarify mathematical models of many probabilistic functions (e.g., sign model, meaning model).

How the robot choose an action given states and feedback? Please clarify the robot policy.

I do not understand why the authors say that inputs by pressing green or red buttons is noisy. It is very clear signal.

What are the 20 pre-defined bases of spoken words?

The description of experimental task should be more clearly. In the used experimental example, how many possible hypothesis of tasks are existing? Please clarify the task and space. Do you consider as It is written that "All results report averages of 20 executions..". What are the 20 executions? Are they considered different tasks: "moving a red cube from location1 to location 3" and " moving a blue cube from location1 to location 3"? Are they different "moving a blue cube from location2 to location 4" and " moving a blue cube from location1 to location 4"?

The explanation of Figure 9 should be improved. What are the tasks of the first and second run? In the first run, what the robot has learned? Please specify how the learned knowledge was used in the second run.

The explanation of the robot execution after learning procedure is missing.

Typo: page 6 "Learning guidance signals: Figure 4" should be "Learning guidance signals: Figure 6"

Comments on the Video Attachment
==========

The video is useful to grasp the concept of the proposed algorithm. The explanation of the robot learning procedure can be improved, for instance, by adding the human voice feedback after each robot action and Showing the clusters in the feature space dur