Robots in the classroom: Learning to be a Good Tutor

Emmanuel Senft[1,*], Séverin Lemaignan[1], Madeleine Bartlett[1], Paul Baxter[2], Tony Belpaeme[1,3]

1 CRNS, University of Plymouth, UK
2 L-CAS, University of Lincoln, UK
3 ID Lab – imec, University of Ghent, Belgium

*** emmanuel.senft@plymouth.ac.uk**

Abstract

To broaden the adoption and be more inclusive, robotic tutors need to tailor their behaviours to their audience. Traditional approaches, such as Bayesian Knowledge Tracing, try to adapt the content of lessons or the difficulty of tasks to the current estimated knowledge of the student. However, these variations only happen in a limited domain, predefined in advance, and are not able to tackle unexpected variation in a student's behaviours. We argue that robot adaptation needs to go beyond variations in preprogrammed behaviours and that robots should in effect learn online how to become better tutors. A study is currently being carried out to evaluate how human supervision can teach a robot to support child learning during an educational game using one implementation of this approach.

1 Introduction

Compared to lectures, tutoring has been showed to increase the learning gains of humans [4]. In particular, one-to-one tutoring enables a more inclusive teaching, by adapting the content of the lesson and the style of interaction to the needs and preferences of the student. As such, tutoring presents numerous opportunities for social robots in education: teaching language [3], how to write [9], maths or sciences [7].

To be as effective as human tutors, robots should not deliver a one-size-fits-all teaching content; they need to adapt their behaviour to the student they are teaching. Traditional methods of developing adaptable robot tutors have either used predefined behaviours that the robot can switch between or have adapted the difficulty of a class to meet the estimated knowledge of the user. But we are convinced that to thrive, robot tutors need to go beyond and learn how to behave efficiently within each situation. Furthermore, we also wish to empower the teachers who are ultimately leading the teaching and who know their students best. Robots should remain tools in the hands of the teachers, and teachers should have the freedom to shape the robot into their own personalised teaching assistant. To this end, we rely on the teacher to demonstrate to the robot the desired tutoring behaviour using a Wizard of Oz (WoZ) approach. As the robot is exposed to these demonstrations, it learns and starts producing its own suggestions of actions to support the students. Using human feedback and commands, the robot's action policy improves over time and when the teacher deems this behaviour to be adequate, the robot can take over the tutoring session, interacting autonomously (if desired) with the students and freeing the teacher to work with other students.

2 Related Work: Adapting Teaching Strategies in Robots for Learning

To increase the amount of learning children gain from the tutoring setup, robots can adapt their behaviour to suit the preferences and requirements of the student they are teaching. One solution, as used in [8], is to have different empathic strategies such as: encouraging comments, scaffolding, offering help or intentionally making errors. By modelling the child's preferences and reactions to these strategies the robot can select the most efficient one for each specific child. Other methods use Bayesian Network and Knowledge Tracing to estimate the learner's knowledge and provide advice on missing skills [10], or select a task and a difficulty level which will maximise the learning gain [6, 11]. Alternatively, if the task requires mainly practice of poor skills (such as handwriting), every aspect of the child's knowledge can be continuously monitored and training examples can be selected to encourage the practice of these poor skills [9].

One method which goes further than simple adaptation and allows the robot to tackle previously unseen or unanticipated child behaviours as a human tutor would, was introduced by Sequeira et al. in [15]. The authors proposed the restrictive-perception Wizard of Oz: the robot starts as non-autonomous; controlled by a human. Then an autonomous controller is developed from the human demonstrations and hand-coded rules before being deployed to interact autonomously and replicate the human demonstrations.

However, in [12] and [13], we argued that the learning of an action policy should occur online, with human supervision. This reduces the workload of the wizard, allowing them to monitor the robot's learning while ensuring that even in the learning phase, the robot's behaviour is efficient. While this method originated from the *Robots in Therapies* field, we are convinced that *Robots in Education* is an area which would greatly benefit from such an approach.

3 Progressive Autonomy for Robots in Education

3.1 A teacher-led learning process

Developed to reduce the workload on a robot's supervisor in a therapy scenario, the Supervised Progressive Autonomous Robot Competencies (SPARC) [12] uses online learning from demonstration combined with suggestions from the robot and potential corrections from the teacher to rapidly learn and improve a robot's action policy.

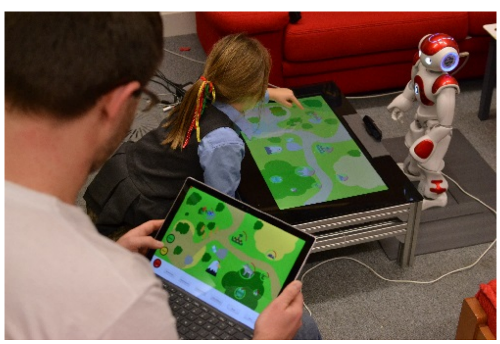

Figure 1. Interaction setup: the teacher (one of the authors) on the left uses a GUI on a tablet to control and teach the robot how to interact with the child until reaching a good action policy.

One advantage of such a technique is that it empowers the end-users, the teachers. They can control the robot's behaviour in a teaching phase, ensuring that the robot reacts properly to the different behaviours expressed by the child while monitoring the progress of the robot's learning (Figure. 1). As the robot learns a better action policy, the teacher can step back and focus more on the child's behaviour while letting the robot progressively take over the tutoring session, freeing the teacher to take care of other children. Keeping the human in the loop and in control of the robot's actions provides the algorithm access to efficient demonstrations and ensures that incorrect actions due to missing knowledge are corrected before being executed, which ensures quick and efficient learning [13]. Having been demonstrated to work only in simple or discrete (in space and in time) environments, this method has not yet been evaluated in a real-world, complex environment such as tutoring.

3.2 A high-dimensional example: a robot tutor to learn about food chains

SPARC for creating a teachable tutor has been implemented in a teaching scenario and is currently being tested with children (source code available [1] [2] [3]).

In this study, children are invited to learn about food chains in a gamified and open learning environment. The setup, as shown in Figure 1, uses the Sandtray paradigm [2] whereby a child is interacting with a robot through a large touchscreen sitting between them. The game presents movable animals and passive plants and the goal is to keep the animals alive as long as possible. Animals have energy that decreases as time goes by and the students have to make them interact with other animals or plants to feed them and replenish their energy. As the students learn how to feed animals to keep their energy high, by extension, they can learn what food each animal eats.

To support this learning, the robot can provide advice (move an animal to, toward or away from other animals or plants), verbal feedback (remind rules, provide congratulation and encouragement) or draw the child's attention to an animal.

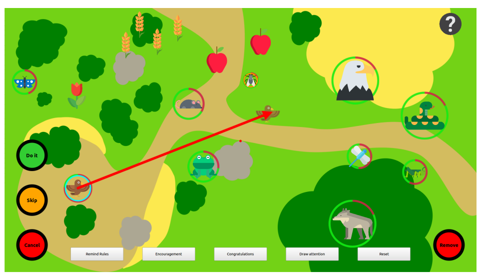

Figure 2. GUI used for supervising: the teacher moves the bird close to the fly and selected both of them as relevant feature for this action (blue and orange circles). Buttons at the bottom are used to have the robot provide feedback.

The teacher uses a tablet running a supervisor GUI replicating the state of the game as it is currently being played on the touchscreen. This GUI allows for remote control of the robot's actions (highlighting features to speed up the learning by providing relevant dimensions for the algorithm [14]) and receives suggestions from the robot about what action to do next (cf. Figure 2).

The robot has access to 655 discrete output actions and an abstracted representation of the state of the game and the interaction through a 210 dimensional vector of values bounded between 0 and 1 (distances between the elements, time since the child touched each elements, time since robot's actions or time since other interaction events). The system must therefore find a correct mapping between a 210 dimensional input vector to a 655 exclusive output one. Many algorithms can learn in such an environment, but traditional Reinforcement Learning algorithms would take a prohibitive amount of time, exhausting many children as the robot would at first be behaving randomly and providing incoherent messages. As such, a method like SPARC offers an opportunity to quickly learn a useful action policy despite the complexity of the environment.

To learn fast, the algorithm used is a variation of the Nearest Neighbours algorithm [5] where actions are defined on a sliced version of the general space [14]. This algorithm allows fast, lightweight and online learning with transparency as the algorithm can highlight which features of the space have been used to make the suggestion.

At the start of the first interaction, the database the algorithm has access to is a blank sheet without any actions, and as the supervisor selects actions, the database of demonstrations is filled, associating actions with the value of the state on a subset of the dimensions. As the database becomes richer, the robot suggests a larger number of correct actions, reducing the workload on the teacher until reaching a point where the teacher only has to correct/select a low number of actions to fine-tune the robot's policy.

This setup is currently being tested in primary schools in the UK with children in

[1] https://github.com/emmanuel-senft/freeplay-sandbox-ros-sparc/tree/task
[2] https://github.com/emmanuel-senft/freeplay-sandbox-qt/tree/food-chain
[3] https://github.com/emmanuel-senft/freeplay-sandbox-qt-supervisor

Years 4 and 5 (8-10 years old) with one of the authors (a PhD student in Psychology naive to the algorithm) acting as a teacher.

4 Discussion

4.1 Future work

The current implementation has several limitations that should be tackled in future work. Firstly, for now, the algorithm can only take demonstrations (and negative feedback) as input. It would be interesting to start with a set of rules defining a baseline of behaviour, which could then be refined online by adding either new rules or demonstrations. Additionally, currently the algorithm only reproduces a demonstrated action policy and does not have the opportunity to learn from its interaction with the world. Future work could focus on designing a system which adds the prediction and use of rewards in reaction to environmental events (such as with Inverse Reinforcement Learning [1]) and techniques to model a child's knowledge to potentially learn an action policy more efficient than the demonstrated one.

While allowing the robot to learn faster using initial knowledge from a human, including a supervisor in the action selection loop also limits the time-scale of the interaction. Allowing the human enough time to correct a suggested action requires the addition of a few seconds between the suggestion of an action and its auto-execution, which implies that the rate of action selection has to be below 1 Hz. This delay can reduce the optimality of an action between its suggestion and execution, slowing down the learning process. Future work could explore teaching at different levels of abstraction, giving the teacher time to override only high level actions where exact timing is less critical.

4.2 Opportunities

The goal of the approach is to provide teachers with a way to create their own personalised robotic tutors, which can be controlled by the teacher and taught how to interact with children according to the teacher's personal preferences. The robot learns from the first demonstration, and to obtain a correct autonomous action policy the teacher would need to spend enough time to cover the required actions in the domain of application. The time dedicated to teach the robot varies with the complexity of the policies from a few minutes for simple ones to more than one hour for complex ones. However, it needs to be pointed out that while the teacher is teaching the robot how to interact, s/he does also actively support students in their learning in a different, while similar, way than traditional human-to-human tutoring.

The mixture between WoZ, learning and autonomy additionally allows the teacher to take a more active supervisory stance for children with more difficulties to offer them an experience tailored to their specific needs, or to select a special (previously taught) action policy for the robot. If the study is successful, we would have demonstrated a way to teach a robot online, an efficient action policy to interact with humans in a complex (high dimensional), indeterministic (children are highly stochastic) environment. This or similar methods could be applied to other domains ranging from personal robotic assistants at home to collaborative manufacturing.

Acknowledgements

This work was supported by the EU FP7 DREAM project (grant no. 611391) and the EU H2020 Marie Sklodowska-Curie Actions project DoRoThy (grant 657227).

References

1. P. Abbeel and A. Y. Ng. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the twenty-first international conference on Machine learning*, page 1. ACM, 2004.

2. P. Baxter, R. Wood, and T. Belpaeme. A touchscreen-based 'Sandtray'to facilitate, mediate and contextualise human-robot social interaction. In *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*, pages 105–106. IEEE, 2012.

3. T. Belpaeme, P. Vogt, R. van den Berghe, K. Bergmann, T. Göksun, M. de Haas, J. Kanero, J. Kennedy, A. C. Küntay, O. Oudgenoeg-Paz, et al. Guidelines for designing social robots as second language tutors. *International Journal of Social Robotics*, 2017.

4. B. S. Bloom. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, 13(6):4–16, 1984.

5. T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

6. G. Gordon and C. Breazeal. Bayesian Active Learning-Based Robot Tutor for Children's Word-Reading Skills. In *AAAI*, pages 1343–1349, 2015.

7. J. Kennedy, P. Baxter, and T. Belpaeme. The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*, pages 67–74. ACM, 2015.

8. I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva. Modelling empathic behaviour in a robotic game companion for children: an ethnographic study in real-world settings. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 367–374. ACM, 2012.

9. S. Lemaignan, A. Jacq, D. Hood, F. Garcia, A. Paiva, and P. Dillenbourg. Learning by teaching a robot: The case of handwriting. *IEEE Robotics & Automation Magazine*, 23(2):56–66, 2016.

10. D. Leyzberg, S. Spaulding, and B. Scassellati. Personalizing robot tutors to individuals' learning differences. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 423–430. ACM, 2014.

11. T. Schodde, K. Bergmann, and S. Kopp. Adaptive robot language tutoring based on bayesian knowledge tracing and predictive decision-making. *Proceedings of ACM/IEEE HRI 2017*, 2017.

12. E. Senft, P. Baxter, J. Kennedy, and T. Belpaeme. SPARC: Supervised Progressively Autonomous Robot Competencies. In *International Conference on Social Robotics*, pages 603–612. Springer, 2015.

13. E. Senft, P. Baxter, J. Kennedy, S. Lemaignan, and T. Belpaeme. Supervised autonomy for online learning in human-robot interaction. *Pattern Recognition Letters*, 2017.

14. E. Senft, S. Lemaignan, P. Baxter, and T. Belpaeme. Toward Supervised Reinforcement Learning with Partial States for Social HRI. 2017.

15. P. Sequeira, P. Alves-Oliveira, T. Ribeiro, E. Di Tullio, S. Petisca, F. S. Melo, G. Castellano, and A. Paiva. Discovering social interaction strategies for robots from restricted-perception Wizard-of-Oz studies. In *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on*, pages 197–204. IEEE, 2016.