SPEAKY Project: Adaptive Tutoring System based on Reinforcement Learning for Driving Exercizes and Analysis in ASD Children

Moussa Nasir, Linda Fellus, Alexandre Pitti

SPEAKY Project: Adaptive Tutoring System based on Reinforcement Learning for Driving Exercizes and Analysis in ASD Children

Moussa Nasir, Linda Fellus and Alexandre Pitti

***Abstract*—Intelligent tutoring systems are increasingly effective for helping the teacher's work with children. However, these technologies are still poorly used for cognitively impaired infants who display autistic spectrum disorders and intellectual disabilities as they don't adapt easily to each infant. We propose an adaptive learning system called SPEAKY for assisting the learning of lexicon to children with the help of the tutor. SPEAKY present a pair of images and questions of gradual difficulty to each infant and adapt the set of images and questions with respect to the child response. Depending on how their tutors scored the child's response, SPEAKY modifies its model of the learner. We proposed an approach based on the reinforcement learning in order to adapt exercises' difficulty to the level and profile of one infant. Our database contains more than 300 images and we have asked more than 2000 questions in three weeks considering all the exercise sessions. The results confirmed that generalization is not possible and that adaptiveness is important as we found that difficulty is child-specific. Through the results we gathered, we also determine the difficulties and facilities points of each child.**

***Index Terms*—**

I. Introduction

Recent works have been conducted on intelligent tutoring systems by [1] using algorithms such as the multi-armed bandit (MAB) in a working environment defined with the help of the educators. In these works, an exploration phase determined the optimum activity for a child according to his profile. The results show that the MAB approach better follows and respects the childs profile in comparison with the more traditional method where the teacher proposed activities matching with infants profile.

L. Fellus is in the Institut Mdico-Educatif du Bois den Haut-APED ESPOIR, 7 rue du Parc, Ennery, France.
E-mail: lindafellus@free.fr

M. Nasir and A. Pitti are in the Laboratory ETIS, UMR CNRS 8051, University Cergy-Pontoise/Paris-Seine, ENSEA, France.
corresponding author: alexandre.pitti@u-cergy.fr

Behind these approaches, the intrinsic motivation of children has been emphasized by [2] to help the learning process and to optimize the learning curve. For instance, when you present some novel stimulus, you could arouse the infants curiosity and let learning to be more effective. In this line, [3] proposed to use an auto-encoder to model the curiosity of children and to study its advantages.

Nonetheless, despite its attractivity, this approach cannot be fully applicable to children having autistic spectrum disorders and intellectual disabilities because of (1) the loose attention they have sometimes and the effort they need to provide to be concentrated, (2) the lack of motivation they get sometimes from just using the system and participating, (3) the lack of a specific purpose of the task if they are not taught by the educator when the questions will finish and (4) MAB algorithms should be tuned as the learner model always changes.

In order to overcome these issues, we propose to develop an intelligent tutoring system called SPEAKY that follows strictly the "Applied Behavior Analysis" (ABA) method [REF] conducted in the medical-educational institute (MEI) du Bois d'en Haut of Ennery (France) where several teams of educators, instructors and psychologists work together with impaired infants. The ABA method consists in analyzing first the behaviour of the child, and then to intervene in order to enhance this behaviour and finally, depending on to the results, the educator chooses a reinforcement or an inhibition for it.

To better understand this approach, we can take into account this example: an educator asks a question to the child, she gives an answer if it is correct, the child receives a reward(a reinforcement), if it is not right then the child's response is corrected and if even the correction is not well assimilated then the infant behaviour is inhibited. This approach is close

to reinforcement learning where you learn the best trajectory to reach a goal thanks to rewards given in each state [4]. Using SPEAKY, we try to refine the children's behaviour by linking good behaviour with good reinforcement, and bad behaviour with inhibition. Although the system is in its first stages, we saw how carefully designed reinforcement algorithms can be effective for children's progression and learning.

In the first part, we will present the environment and the implementation of our system, the second part will be devoted to the protocol we set for conducting the exercises and finally in the third part, we will discuss about the results.

II. Material and Methods

A. Experimental Setup

After several meetings organised with the MEI, we have jointly decided that the proposed tutoring system will be in the form of visual questions and oral answers. An image is presented to the child that he will try to speak about its content to the tutor. The images are part of eight different lexical themes (nature, beverage, fruit & vegetable (F&V), food, school supplies, daily objects, vehicles, animals) and they can be presented with seven different styles (icon & sketch, pictogram, pictogram in black & white, photo context, photo context in black & white, uncluttered photo, uncluttered photo in black & white).

An exercise in SPEAKY takes place as follows within the framework of the ABA method: a question is presented, the child gives an answer, the child's teacher carries out a verbal reinforcement and then gives a quotation (a score). When the exercise is finished, the tutor gives the quotation to the system for adapting the algorithm for the next sets of questions. The purpose of ABA is also to explain to the child why he should do this exercise, in our case we explain to him that making an exercise will lead him to a reinforcement. However, in some cases, there is no need to talk about a reward because some children are very curious about just participating in an exercise but not all.

The reinforcers vary according to the children. They are of different kinds and can be graded depending on the value which attributes each child. For the first ones, the educator congratulates and encourages the child, in some cases after a few questions, the educator can give a candy to the child in order to keep his motivation. For the final reinforcer, as all exercises were done before going to lunch, the general reinforcer was to eat at the canteen. Other final reinforcers were also used, for instance, listening to a song, playing a game on a tablet, offering him a drawing to colour, playing with modelling clay. Reinforcers were chosen by educators who know the children well enough.

The tricky part of this exercise was the choice of the quotation to give for a given answer by the child. The idea was to have an objective, universal quotation that best reflected the level of the infants. We also asked ourselves the question of what is really evaluated. Indeed, in our case, we will evaluate the understanding of the image, we do not evaluate the child's pronunciation as long as it does not prevent us from understanding the child's response. And since the goal is to increase the children's vocabulary, we will focus on learning word about the topics chosen previously.

We chose to set 4 possible ratings for a given answer. A score of "0" will be given if the child did not understand the question as well as the exercise, a score of "1", if he makes a wrong answer. He will get a score of "2" if his answer is in relation to the image. Finally, a score of "3" will be given if the answer perfectly describes the image.

Fig. 1. Screen display of the epured HMI. At each trial, one image is presented randomly selected with a question; here "what is this?" in french. The images are part of eight different lexical themes (nature, beverage, fruit & vegetable (F&V), food, school supplies, daily objects, vehicles, animals) and they can be presented with seven different styles (icon & sketch, pictogram, pictogram in black & white, photo context, photo context in black & white, uncluttered photo, uncluttered photo in black & white); see Fig. 2.

B. Implementation

SPEAKY used the Pygame library in python language to build the HMI presenting the questions.

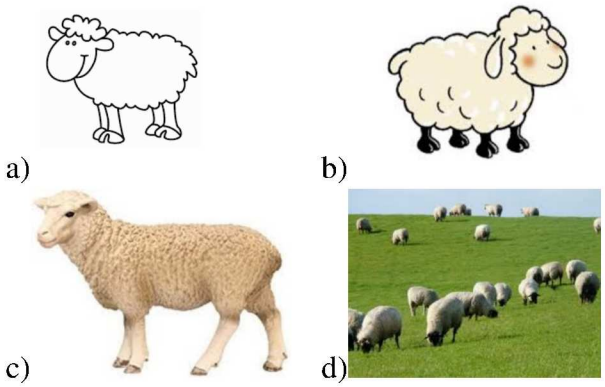

Fig. 2. One example of an image presented with different styles; resp. pictogram, sketch, uncluttered, color photo in context.

Our database gathers all the images classified by theme and style as discussed in the previous part. It consists on 8 subjects and each subject has an average of 6 to 8 images. In addition, each image can be represented in 7 different ways (i.e., styles). Therefore there is between 336 to 448 images.

The idea is to use a reinforcement learning algorithm to try to estimate the best trajectory toward a difficulty. For example, if a child has trouble identifying an eraser, we will first offer him pictures of a school bag, a pencil case and a pencil so that the child can build his categorization by associating images from a common theme in order to reach a difficulty. We want the child to work on his associative memory. Nevertheless, we also want the infant to be flexible, in other words he does not have to be totally focused on a theme. In fact, he should be able to juggle between the subjects while trying to be effective. Finally, the dilemma between associativity and flexibility can be compared to the exploitation and exploration dilemna in the reinforcement approach. We will explain this comparison in this part.

C. Q-learning Reinforcement Algorithm

We used an algorithm quite close to the Q-learning, but there are some lightly differences that we will try to show. We can make the analogy with a reinforcement algorithm used in video games since our goal is to win against the child by asking more or less difficult questions. The general working of this algorithm is to perform an action in our case it is to "ask a question", the notation of the answer is considered as the reward of this action. A matrix of the reinforcers is thus updated, whose themes are in x and the styles are in y. Therefore, we have an 8x7 matrix with 56 states with each state corresponding to images characterized by a theme and a style. Likewise, for each action, a value function is computed, so there is also the matrix of the value function, which has the same dimensions as the reward matrix and is updated after each action.

The value function is updated as follows:

$$\Delta Q = Q(s, t+1) - Q(s, t) \tag{1}$$

$$\Delta Q = \alpha * [R(s,t) + \gamma * \min\left(Q(s', t)\right) - (1-\gamma) * Q(s,t)] \tag{2}$$

with s the current state, s' the next state, R the reward (the quotation given by the educator), α the learning rate and γ the discount factor.

For the next state, we take the action whose value function is the lowest. Indeed, in our matrix of the value function, the states with the lowest values correspond to "not visited" states or "difficult" states (less right answer). And the algorithm must target the difficulties of the child, this targeting is done more or less gradually depending on the profile of children, for this we will play with several parameters.

1) Random exploration to Greedy search: The learning rate allows to manage the speed of convergence of a state towards an "acquired" state or a "difficult" state. In other words, according to its value, we will consider an "acquired" state after a certain number (depending on the parameter) of good answers and on the contrary, we will consider a "difficult" state after a number of wrong answers. For instance, if we take a high learning parameter, it will take 1 or 2 correct responses in a state for its value function to converge to a threshold that would tell us that the state is "acquired". Likewise, if the learning rate is weak, it will require many more good answers before a state converges to an "acquired" state. The example applies in the same way for wrong answers that will bring a state to a state considered "difficult". In addition, the discount factor allows us to manage the recent action influence compared to the old actions.

As we explained in the introduction of this part, there is one essential parameter to manage, this is the exploration parameter ϵ. Like the learning parameter, its value will be entirely conditioned by the child's profile. Nevertheless, its value is updated after each action, we multiply it by

a coefficient between 0.9 and 0.99 in order to minimize it progressively. The idea is to explore the states on the first 20 to 30 questions and then exploit for the rest of the questions (10 to 20). However, the exploration and exploitation presented in this part are valid at the general scale. In fact, we take into account the entire matrix of states without really considering specifically the theme and style of images; hence, we can qualify this approach as a first-level approach (intersystem).

2) Variational Q values search: We used a second level of exploration and exploitation, which is managed according to the variation in the value function:

$$\delta Q = |Qcurrent - Qprevious| \tag{3}$$

This second level permits to manage exactly the questions that are too easy or too difficult. For instance, if the child faces 2 easy questions (2 correct answers), the variation will be zero, it means that is necessary to explore, in other words change the subject and the style. The example applies in the same way for 2 difficult questions (2 wrong answers). Now, if the variation is not zero, we can estimate that there has been either a positive improvement where the child learned, or it could be a negative progression where the child has moved from an easy question to a difficult one and did not know how to answer the difficult question. In its 2 cases we fixed the theme and we change the style.

Obviously, we set a threshold for the value function variation, according to which the progression is significant. Moreover, we can consider several Q previous value to best estimate the variation. This second level (intra-system) is essential to adapt the difficulty and to build vocabulary word learning by playing on themes or styles.

The algorithm works in real time, there is no ”pre-learning” phase where we usually use a random policy. Both matrices (rewards and value function) are initialized to 0 for the first work session. Then for the other sessions, it is obviously necessary to save and reuse the matrix of the value function that has been learned.

III. Results

At first, we test our algorithm using the Random-to-Greedy-Search policy. We took α = 0.5, γ = 0.1 and the exploration parameter ϵ= 0.9 (ϵ is decreased by 0.95 in each iteration).

We used this algorithm in the last week of work sessions with infants. A01, J01, B04 and L05 have participated in these sessions. We present the results of A01 and B04 between two sessions, respectively in Figs. 3 and 4. We know that A01 has a better level than B04. Therefore, we choose a learning parameter α at 0.5 for A01 and 0.2 for B04. In this way, the algorithm required better answers to validate a state as being acquired. Moreover, we set a better exploration parameter for A01 because we supposed he would better use his flexibility than B04. We asked 50 questions for each session and here the results for two sessions.

The Q values of A01 are displayed in Fig. 3 between two sessions. We can see that A01 better performs with uncluttered photo. He improves himself in the state (grayscale pictogram; Nature) between the two sessions. We noticed that when the child reaches highest score in a theme, then the probability that the algorithm proposes the same theme with a different style is lightly low. The algorithm may consider that the child mastered the theme so basically that it is better to select another theme

The Q values of B04 are displayed in Fig. 4 between two sessions. In comparison to A01, B04 has been in difficulty with the sketches and pictograms (for F&V, Supplies, Beverage) but the algorithm proposed the same themes in different styles (Context photo, Grayscale uncluttered photo) and then he succeeded in these difficult subjects. The use of a different learning paramter α than A01 makes the algorithm to perform differently, forcing more exploration.

In a second part, we tested the second policy with variational Q values and compared its results with the first policy presented earlier, see Fig. 5. In order to compute the variational policy, we computed the difference between the Q values matrix taken between two sessions. We found that the results were better for this policy than for the first one with infants a little bit more concentrated with the cumulated quotation level always increasing and not decreasing as it was the case for the first policy.

IV. Discussion

We proposed an adapative tutoring system based on reinforcement learning and ABA method for

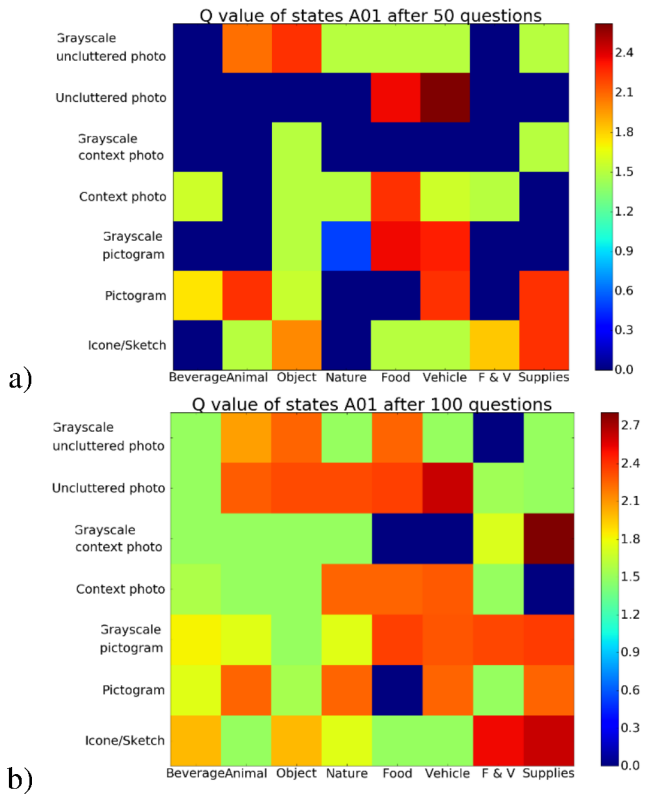

Fig. 3. Q values comparison between two sessions and after 50 questions; respectively a) and b). We can see an overall exploration of the different questions so that the algorithm proposes the same theme with a different style (vertical lines) is lightly low.

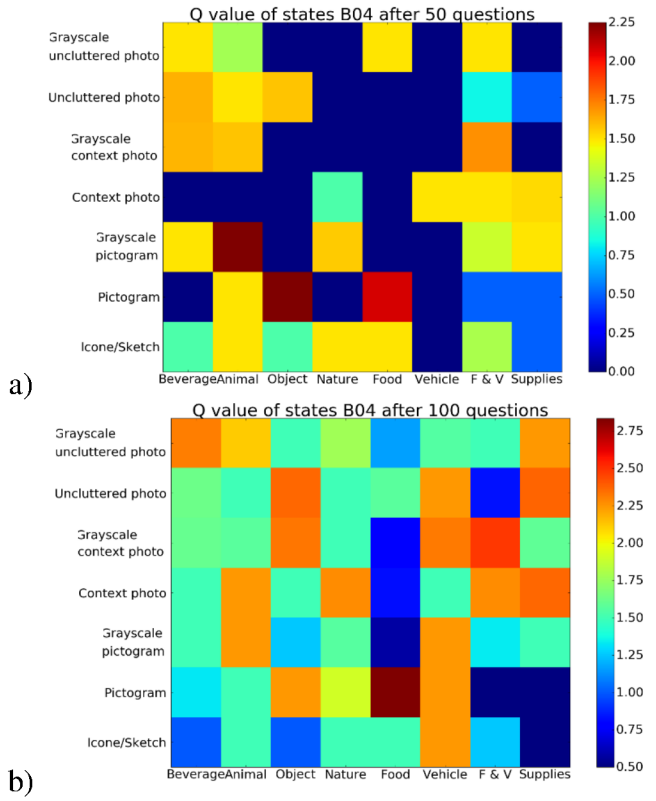

Fig. 4. Q values comparison between two sessions and after 50 questions; respectively a) and b). see text for explanations.

infants with broad cognitive syndrome disorder. Although, in its primarily stage, our results might be not representative enough as it has been used during 3 weeks with only one group of infants. Therefore, these results may be difficult to compare and interpret because the infants' level was better in the last week than in the first week. This is nevertheless one known problem in evaluating intelligent tutoring systems in general and protocols have to be adapted cases by cases.

In our approach, for instance, the issue was to set the correct parameters for each children before each session. If we set a too high learning rate, the algorithm would consider a state as acquired too fast. In the second test sessions of Q learning, we did not use the first level of exploration (see section II-C) because we consider that we get enough information with the first test sessions. We only exploit and use the second level of exploration only if a state is too easy or too difficult. As shown in the Fig. 5, the results are slightly better in the second session using the variational Q-learning.

In future work, we think possible to improve the reinforcement approach. For instance, we can think to have a neural network that can update the Q-learning parameters to choose the best learning rate and exploration parameter, considering infants profile and his progression. In all these approaches, our states had been characterised by selecting the theme and the style of image database. We should try to consider in future works widen the number of images in our database. For example, we have only 6 or 7 different images of pictograms and animals, which could be considered in the Q matrix.

ACKNOWLEDGMENTS

This work was supported by the Institut Medico-Educatif du Bois den Haut, Ennery France, and APED ESPOIR, and the ETIS laboratory. We thank Chloé Bryche from the Institute for reviewing the paper and students M. Koreissi, R. Dinar, E. Fouache and K. Furet for developing a previous version of the Speaky software.

REFERENCES

[1] B. Clement, D. Roy, P. Oudeyer, and M. Lopes, "Multi-armed bandit for intelligent tutoring system," *Journal of Educational Data Mining*, vol. 7, no. 2, pp. 20–48, 2015.

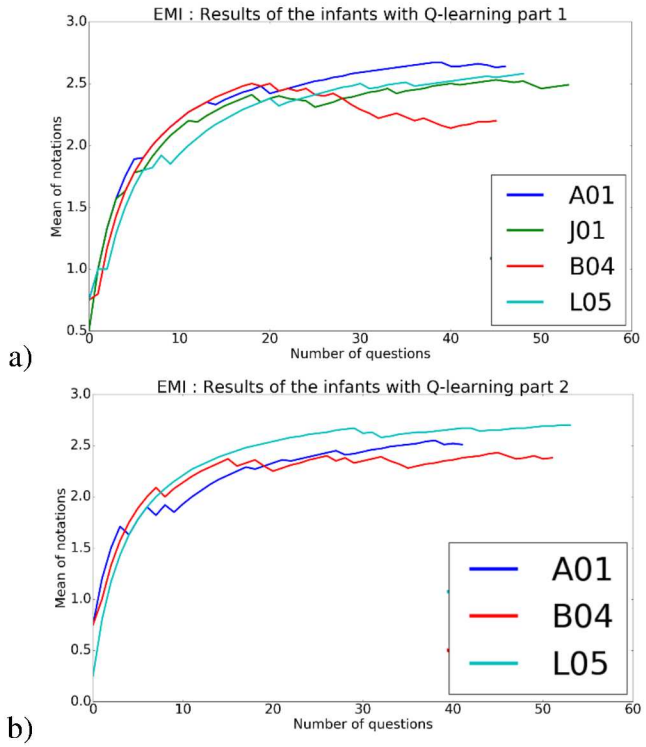

Fig. 5. The results are better in the second part. We used in the second part the Q value matrix get at the first part.

[2] P. Oudeyer, J. Gottlieb, and M. Lopes, "Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies," *Progress in Brain Research*, vol. 229, pp. 257–284, 2016.

[3] K. Twomey and G. Westermann, "Curiosity-based learning in infants: A neurocomputational approach," *Developmental Science*, p. e12629, 2017.

[4] A. Barto and R. Sutton, "Reinforcement learning in artificial intelligence," *Advances in Psychology*, vol. 121, pp. 358–386, 1997.