An Imitation Game for Emerging Action Categories

Bart Jansen

Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium,
bartj@arti.vub.ac.be
http://arti.vub.ac.be/~bartj

Abstract. An imitation game for learning action categories is proposed. It serves to invent and share a repertoire of action categories in a population of robotic agents. The agents start without any action categories built in, but they learn by imitating other agents and gradually invent categories for the actions they observe and execute. In other words, no action categories need to be defined beforehand, nor do agents have to be assigned fixed roles. If the population only consists of two agents, the repertoire of action categories can be learnt without feedback about the outcome of the game between both agents.

1 Introduction

Most of the work that has been published on learning by imitation in robotic agents focuses on the learning of action categories in a teacher - student context, either the student learns from other robots [1] or from human beings [2,3,4]. In such a set-up one agent acting as teacher already has action categories. By observing the teacher who is executing actions, the action categories can be transferred to the student: by imitating the teacher's action using for instance inverse kinematics and evaluating that action, the learner can know whether it correctly reconstructed the observed action. However, such a set-up does not explain how imitable action categories emerge and does not take into account the influence of the population dynamics of imitating agents. The emergence of categories on the population level has been studied intensively in our laboratory, for example the emergence of vowel systems [5], colour categories [6] and visual categories [7].

In this paper we propose a learning mechanism for agents such that new action categories can emerge when imitation of actions fails. This is done in a population of agents engaging in imitative interactions, called *imitation games*. The learnt action categories will be imitable. If an action is hard to observe or to imitate, it will not be learnt by other agents. The concept of imitation games used here strongly resembles the concept of imitation games used in [5] in the context of vowel systems.

Section two briefly explains the experimental set-up and the architecture of the agents. Section three describes the imitation game in detail. Section four

W. Banzhaf et al. (Eds.): ECAL 2003, LNAI 2801, pp. 800–809, 2003.

contains results showing that a repertoire of action categories can indeed emerge in a population of agents through imitation games. In section 5 a modified imitation game is presented. Playing that game, two agents can acquire a repertoire of action categories without giving feedback on each other's imitative behaviour. Section 6 discusses future work.

2 Experimental Set-up

An agent has a head with a stereo camera and a robot arm with a gripper. Using its arm, the agent can perform different kinds of gestures which it can observe using the vision system. We have built this physical set-up using a readily available commercial robot arm, called Teach-robot[1] and a vision system based on the Small Vision System (SVS)[2]. The vision system focuses on finding the coordinates of the gripper in the captured images. The resulting three-dimensional time series will be categorised during the imitation game. The vision system and the robot arm are described in more detail in [8].

The experiments described in this paper are not performed on this real physical set-up, but in simulation and serve as a proof of concept. Action execution and observation are thus not performed on real robots but are simulated using the forward and inverse kinematics of the robots. Both in the simulated execution and observation of actions, noise (e.g. 5%) is added for simulating imprecision when performed on real physical robots.

2.1 The Agent Architecture

An *action category* consists of two components: an *action* and an *observation*. This means that every category is represented by a single prototype. An action is a tuple of two 3D points, being the start- and endpoint of the gripper. Gripper points define an *action space*. The action executed by the robot will be the movement of the gripper from start point to end point. The observations are sequences of 3D points, generated by the vision system. They define an *observation space*.

Actions do not contain direct motor commands, but tuples of gripper coordinates. There is a direct mapping between both, as the forward and inverse kinematics of the agent's robot arm are known [9]. Given an action, the agent can obtain the associated observation, simply by executing the action and observing the result. Given the observation, the agent can only recover the action that was executed if the action space and the observation space are calibrated. In section 5 it is discussed whether calibration is indeed essential for the imitation game.

Agents have an *action category memory* to store the learnt action categories. At the beginning of the imitation games, these memories are empty as the action categories will emerge during the imitation games. Besides a memory for

1 Microelectronic Kalms, http://www.teach-robot.de/
2 http://www.ai.sri. com/konolige/svs/Papers

the learnt action categories, all agents are equipped with the same learning mechanism and have a pre-programmed drive to act and to imitate. We do not investigate how the agents can learn how to imitate nor when to imitate. We rather start with a built in learning mechanism and investigate whether this mechanism is suited for building up a shared repertoire of action categories.

3 The Imitation Game

For every game two agents are randomly selected from the population. One agent will take the role of *initiator*, the other agent will be the *imitator*. During the game, agents will observe the execution of actions and will try to categorize their observations. The category of an observation O is defined as the action category in the agents memory with an observation closest to the observation O. We use Dynamic Time Warping (DTW) [10] as distance metric on observations of actions. DTW was also used for categorising gestures in [11]. The distance metric that is used by the agents for comparing observations defines the type of categories that will emerge. If the distance metric only classifies two observations as equal when the entire trajectories match, every consistently repeatable and observable motion of the robot arm will constitute a different action category. This explains why so many categories will emerge.

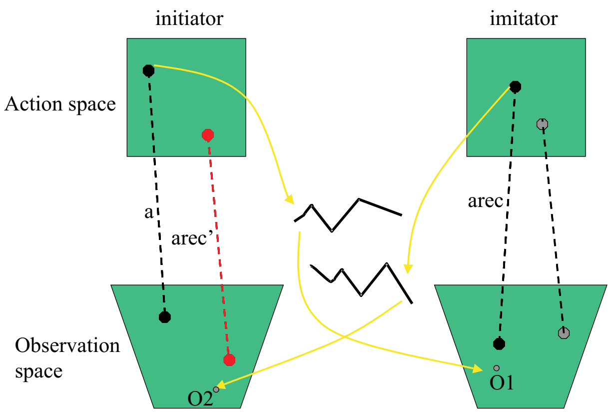

Fig. 1. The imitation game. In this example it fails, as the imitation is categorised as a'_{rec} instead of a.

In figure 1 the observation space for both the initiator and the imitator are depicted. The dashed lines represent the associations between actions and observations, which constitute categories together. Those observations and actions are indicated by large dots, the small dots indicate the actual observations.

If both the initiator and the imitator have at least one action category in their memories, they can engage in an imitation game. Otherwise, they first add a new random action category to their action category memory. The initiator then selects a random action category a from its memory and executes the action that is associated with the action category a. The imitator observes this action, denoted *O1*. It categorises the observation *O1*, denoted a_{rec}. The imitator now executes the action associated with the category a_{rec}. This action will be observed by the initiator as observation *O2* and categorized. The category will be called a'_{rec}. If the initial category and the category of the imitated action are equal ($a = a'_{rec}$), then the initiator decides that the game succeeds, otherwise that it fails. The initiator sends non-verbal feedback about the outcome of the game to the imitator. If the game succeeded, the imitator knows that the category it used is similar to the category used by the initiator. To increase similarity, the imitator shifts the category it used closer to the observation *O1* of the initiators' action. If the game fails, two different update strategies are considered. If the action category the imitator used has been successful in the past (i.e. its success-ratio is above 0.5), the failure in this game is probably caused because the initiator executed an action the imitator does not know yet. In this case a new category is constructed for the observation. If the success-ratio of the category used is below 0.5, the category itself is probably not very good. In that case, the imitator shifts the action category used towards the observation, to improve the success of the category in the future. As a last step of the game, both the initiator and the imitator update their repertoire of action categories. If an agent has an action category of which the action was not successfully imitated, the action category is removed from its repertoire (i.e. after five or more categorizations and when the success-ratio is below 0.7). Agents are forced to invent and share new action categories because with a small probability (*e.g. 0.02*) new random action categories are added to their repertoire.

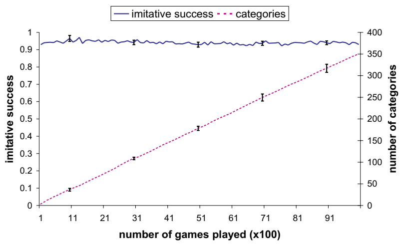

Fig. 2. Number of categories and imitative success averaged per 100 games over 10000 games, for 2 agents. Results are averaged over 10 runs.

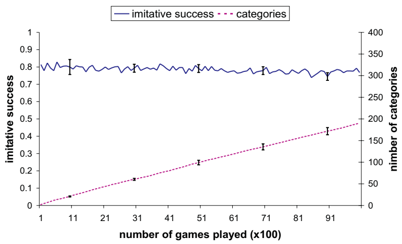

Fig. 3. Number of categories and imitative success averaged per 100 games over 10000 games, for 5 agents. Results are averaged over 10 runs.

4 Results

In the first experiment the population consists of 2 agents. In total 10000 imitation games were played. In figure 2, the imitation success is shown averaged per 100 games, averaged over the 2 agents, averaged over 10 runs. The average number of action categories per 100 games averaged over both agents and over 10 runs is shown in the same figure. The number of action categories rises steadily and reaches (345.46 ± 2.34) after 10000 games. Even while the number of action categories increases, the imitation success remains very high (0.95 ± 0.01).

These results indicate that two agents indeed build up a shared repertoire of action categories through imitation. Averaged over the 10000 games, in 95% of the cases their imitation is successful. In larger populations, agents also build a repertoire of action categories, but the imitation success is lower. In figure 3 the average number of categories and the imitation success are shown for 10000 interactions among 5 agents. The results are averaged over 10 runs. Averaged over the 10000 games, the imitation success is 0.80, which is much lower than when only two agents were used, but still higher than in the case of interactions between agents with random categories. It is not surprising that the imitative success is already very high at the beginning. As long as all agents have only one action category, there is no possibility of failure. Further experimentation is required to investigate how these results depend on the different parameters used in the game. Particularly, it needs to be shown that this game is robust enough to deal with larger amounts of noise.

Measuring category similarity. From the results shown before, it is obvious that the agents are very successful in imitating each other. Although it is easy to understand that successful imitation is not possible if the agents have very different action categories, it must be shown that the categories are shared through the population. Therefore we need to define how similar the categories of two agents are, or—in general—how similar two sets of points are in an N-dimensional space. Note that two agents can develop a different number of categories, so the sets do not contain an equal number of points. The similarity measure of the categories of two agents *Category Distance*(CD) is based on the *Weighted sum of minimum distances* metric developed in [6] and is given in equation 1. The terms A and B are agents, a and b are categories and $d(a, b)$ is the distance between the categories a and b.

$$CD(A,B) = \frac{\sum_{a \in A} \min_{b \in B} d(a,b) + \sum_{b \in B} \min_{a \in A} d(a,b)}{|A| + |B|} \tag{1}$$

We can now calculate the Category Variance (see equation 2) of the population of agents. This measure indicates how much the categories of all agents deviate from each other.

$$CV(population) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} CD(A_i, A_j) \tag{2}$$

As opposed to the imitation success and the number of categories, the category variance is not averaged over 100 games. In order to reduce computation time, it was only calculated every 100 games. In figures 4 and 5, the category variances for two and five agents playing 10000 imitation games are shown. It can be seen that the category variance decreases very fast in both games, so the more imitation games the agents play, the more similar their categories become. However, the reader should notice that the CV also decreases in a population of agents with an increasing number of random categories, simply because more categories are fitted into a finite space and thus become closer to each other. In figures 4 and 5, one can see that the CV is always lower in the case of agents that learn through imitation than in the case of agents with an equally fast growing number of random categories (ECV).

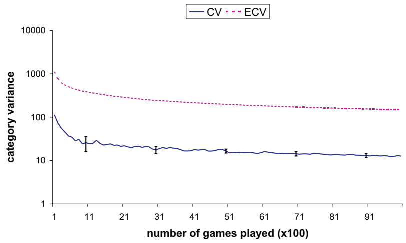

Fig. 4. The category variance for 10000 interactions among 2 agents, in a logaritmic plot. Results are averaged over 10 runs.

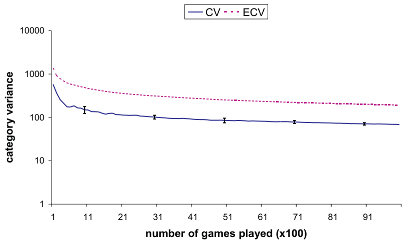

Fig. 5. The category variance for 10000 interactions among 5 agents, in a logaritmic plot. Results are averaged over 10 runs.

5 A Variation on the Game

In this section it is investigated how the paradigm of imitation games presented above can be modified such that there is no feedback required between both agents.

5.1 Who Is Learning?

In the imitation game as described above, one could ask who is learning—the initiator or the imitator? Both initiator and imitator are keeping success scores, removing bad actions and randomly adding new actions. However, most important is that the imitator considers the initiator as an agent with perfect categorization and adapts its categories by shifting to resemble the categories of the initiator more closely. As all agents take turns and take many times the role of initiator and imitator, the result is that all agents' categories become more similar. Another perspective is also possible. Suppose that the initiator considers the imitator to have full knowledge of the categories. In that case, the initiator uses the imitation to check whether its own categorization of the world is consistent with the categorization of the imitator. If the initiator's action category and the categorization of the imitation are equal, the initiator is ensured that its own categorization (for that single action) is very similar to the categorization of the imitator. So, the initiator adapts its categories to resemble the imitated action even more . If the initiators' action category and the categorization of the imitation are different, the initiator adapts its categories to decrease the probability of failure in the future. What is the difference between both perspectives? In the first game, the agent which learns by adapting its categories (the imitator) receives non-verbal feedback about the outcome of the game and adapts its categories accordingly. In the second game, there is no non-verbal feedback, as the initiator itself knows the desired and actual outcome of the game. An important difference is that in the new game, the agent that is learning—the initiator—can choose what action categories to use in the game. The initiator can investigate its own repertoire of action categories for categories that were not successfully imitable in the past. By using precisely those categories, the agent might improve its unsuccessful categories very fast. Another possibility is that the agent always selects categories with a high success ratio. Further research on this topic is required. The second game in which the initiator learns action categories is presented below.

5.2 The Learning Initiator

A modified game is proposed in which the initiator and not the imitator is learning. In this game, the initiator selects a random action category a from its memory and executes it. The imitator observes it as *O1* and categorizes it as a_{rec}. The imitator executes the associated action, which is observed by the initiator as *O2* and categorized as a'_{rec}. If a and a'_{rec} are equal, the game succeeds, otherwise it fails. So far, the game is identical to the first game. At this stage of the game the initiator does not send feedback about the outcome of the game to the imitator, as the initiator updates its action categories and the imitator does not. The initiator shifts its original action category a towards the action category a'_{rec} of the observed imitation. The shift operation does not start from the observation space but is done in the action space. In case of a failure, this same shift can be done in case of an action category a with high success. If the action category a does not have high success, it is not useful to add a new action category for the observation associated with a'_{rec}, as this will result in a new action category a''_{rec} which is very close to a'_{rec}. This addition would only confuse the agent. Therefore in this case, no new action category is added.

Shifting in the action space instead of the observation space only works if it is guaranteed that when an action is shifted towards another action, then the observation of the first action is also closer to the observation of the second one. This is the case in a real robot, as points in both spaces are related by a rotation and a translation.

The advantage of shifting directly in the action space is that the agent must never calculate what action caused a given observation. This means that the action space and the observation space need not to be calibrated anymore. This has an enormous advantage for performing experiments on real world robots, where calibration between observation space and action space would not be required anymore.

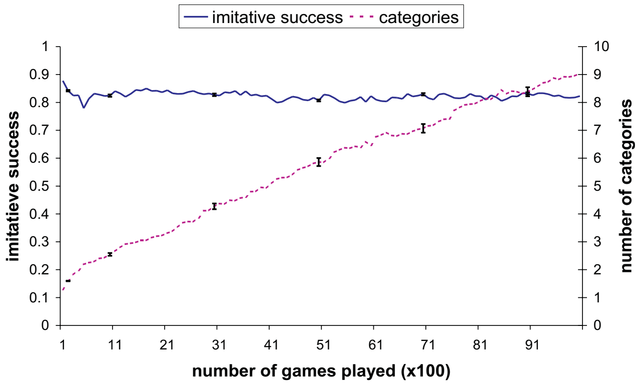

Fig. 6. Number of categories developed over 10000 games where the initiatator learns, averaged per 100 games. The initiator is adapting its repertoire. Imitative success averaged per 100 games, for the same 10000 games is shown in the same graph. Results averaged over 100 runs. The action space and the observation space are not calibrated.

Results are given in figure 6. Although convergence is slower than in the previous type of game, the agents are capable of developing a repertoire consisting of 9.04 ± 0.89 action categories, while the imitative success remains (0.82 ± 0.01). At the moment, it has not been possible to show that a population of more than 2 agents is capable of converging to a shared repertoire of action categories by playing the imitation game presented in this section. Using the same parameters as in the former imitation game, the average number of categories for a population of 5 agents after 10000 interactions was found to be 1.28 ± 0.14. Other parameter settings have not been investigated yet. The fact that the repertoires are built up much slower in this type of game and the fact that imitation games are only successful in populations of two agents is caused by the lack of feedback and by the fact that the initiator does not adapt its categories to become more similar to the categories of the imitator.

6 Future Work

This paper has shown how a repertoire of very simple action categories can be learnt and shared throughout a population of agents. However, the question remains how more complex actions can be learnt. We believe that this is possible by using the action categories as they emerge in the imitation game as building blocks for learning more complex actions. On a higher level, a new imitation game could be defined, such that complex actions can be learnt. The bigger challenge is developing an imitation game where actions involving object manipulations emerge, without the manipulation of the objects being predefined, but as an emergence of the task of imitation and a utility function.

7 Conclusion

In this paper it was shown that agents can learn action categories through imitation, even if no strict teacher-student protocol is followed where the teacher already has fully developed action categories. The agents start without any action categories and invent a repertoire of action categories while imitating other agents from the population. Due to pressure of the game to improve categorization, the action categories become shared among the population and can thus be imitated with high success (above 90%) by the other agents. This is not guaranteed if the action categories were pre-programmed. This game is also working in larger populations. The obtained results serve as a proof of concept and encourage us to repeat the same experiments on the robot set-up that we have built.

If the population is restricted to only two agents, we show that the agents can develop and share a repertoire of action categories without feedback about the outcome of the game between both agents and without the agents' action space and observation space being calibrated. This is an enormous advantage for conducting experiments on real robots. To achieve this we proposed a second imitation game in which the initiator uses the imitated action of the imitator to update its categories.

Acknowledgments. Bart Jansen is sponsored by a grant from the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT).

References

1. Vogt, P.: Grounding language about actions: Mobile robots playing follow me games. In Meyer, Bertholz, Floreano, Roitblat, Wilson, eds.: SAB2000 Proceedings Supplement Book, International Society for Adaptive Behavior (2000)
2. Billard, A., Schaal, S.: Robust learning of arm trajectories through human demonstration. In: Proceedings of the International Conference on Intelligent and Robotics Systems, IEEE Computer Society (2001) 82–90 Hawaii, November 2001.
3. Billard, A., Hayes, G.: Transmitting communication skills through imitation in autonomous robots. In Birk, A., Demiris, Y., eds.: Proceedings of EWLR97, Sixth European Workshop on Learning Robots, Lecture Notes on Artificial Intelligence. Volume 1545., Springer (1997) 79–95 Brighton, UK, July 1997.

4. Alissandrakis, A., Nehaniv, C.L., Dautenhahn, K.: Do as I do: Correspondences across different robotic embodiments. In Polani, D., Kim, J., Martinetz, T., eds.: Proceedings of the Fifth German Workshop on Artificial Life (GWAL5). (2002) 143–152 18-20 March 2002, Lübeck, Germany.
5. de Boer, B.: Self organization in vowel systems. Journal of Phonetics **28** (2000) 441–465
6. Belpaeme, T.: Factors influencing the origins of colour categories. PhD thesis, Artificial Intelligence Lab, Vrije Universiteit Brussel (2002)
7. Belpaeme, T., Steels, L., Van Looveren, J.: The construction and acquisition of visual categories. In Birk, A., Demiris, Y., eds.: Proceedings of EWLR97, Sixth European Workshop on Learning Robots, Lecture Notes on Artificial Intelligence. Volume 1545, Springer (1998) 1–12 Brighton, UK, July 1997.
8. Jansen, B., de Vylder, B., de Boer, B., Belpaeme, T.: Emerging shared action categories in robotic agents through imitation. In Dautenhahn, K., Nehaniv., C.L., eds.: Proceedings of the Second International Symposium on Imitation in Animals and Artifacts., University of Wales, Aberystwyth (2003) 145–152
9. De Vylder, B.: Forward and inverse kinematics of the teach-robot. Technical Report AI MEMO 02-02, Artificial Intelligence Lab, Vrije Universiteit Brussel, Brussels (2002)
10. Myers, C.S., Rabiner, L.R.: A comparative study of several dynamic time-warping algorithms for connected word recognition. The Bell System Technical Journal **60** (1981) 1389–1409
11. Corradini, A.: Dynamic time warping for off-line recognition of a small gesture vocabulary. In: IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS'01), IEEE Computer Society (2001) 82–90 Vancouver, B.C., Canada, July 13–August 13, 2001.