Imitation of actions between differently embodied agents, without kinematics models and without feedback

Bart Jansen
Artificial Intelligence Lab
Vrije Universiteit Brussel
Pleinlaan 2
BE - 1050 Brussels
bartj@arti.vub.ac.be

Abstract

We introduce the paradigm of imitation games and show how repeated local interactions between agents lead to a shared repertoire of actions. Three of the most restricting assumptions which were made in previous work are analyzed and discussed and shown to be unnecessary. By relaxing the initial assumptions, imitation is possible without feedback between the agents, agents can have different embodiments and agents do no longer require a model of the kinematics of their body.

In this paper, all three assumptions stated here will be investigated and it will be shown that none of them is actually required. By relaxing those three assumptions, the flexibility and versatility of our imitative framework is illustrated.

2. The Imitation game

In our work, every individual is represented as an agent. Several agents exist in a population. Within this population, no agent starts with a set of actions, i.e. all agents start tabula rasa. The agents interact with each other and by interacting each agent gradually constructs a repertoire of actions. Both in our computer simulations and with real physical robots, we have shown how these repeated local interactions can lead to a shared set of actions which can be observed and imitated by all agents. As such, the global sharedness of the learnt actions is an emergent phenomenon (Jansen, 2003b).

A single imitation game is an interaction between only two agents. Those two agents are selected randomly from the population. One of them takes the role of initiator, the other agent takes the role of imitator. The game is played as follows:

1. The initiator randomly selects an action $a'$ from its repertoire and executes this action. If its repertoire is empty, a new random action is first added.

2. The imitator observes the action, finds the best matching action $a''$ from its own repertoire (categorial perception) and executes the action. If its repertoire is empty, a new random action is first added.

3. The initiator observes this action, finds the best matching action $a''$ from its own repertoire and compares its initial action $a$ with the recognized imitated action $a''$.

4. If both actions are the same, the game succeeds, otherwise it fails.

1. Introduction

In previous work we have introduced the framework of imitation games as a powerful mechanism for the learning of shared repertoires of actions. This approach has several important characteristics. Opposed to many other work on robot imitation, we consider agents which all start without any built in actions. By repeated local interactions dubbed imitation games, the agents gradually invent a shared repertoire of actions. So, there is no initial teacher from which all actions are passed to the other agents. In our approach, all agents can take the role of initiator and imitator and those roles are assigned randomly during the interactions.

Although the emergence of the shared repertoires is a merit of our imitation framework, the framework is based on several strong and unmotivated assumptions. Among others, a perfect single bit communication channel is supposed to exist between both agents participating in the interaction. The channel is used for feedback, which is required for the proposed learning mechanism. Besides the feedback, it is assumed that all agents have the same embodiment and that all agents have a model of their forward and inverse kinematics. This latter assumption facilitates learning heavily: with a good model of its own kinematics, it is easier to learn to reproduce an observed act.
5. The initiator announces the outcome of the game to the imitator.

6. The imitator adapts its repertoire using this feedback, such that future games become more successful.

7. Both agents adapt their repertoires, independent of the current game.

On an abstract level, the game thus consists of three consecutive steps: interaction(1–4), sending feedback(5) and learning(6–7). Since we do not study learning to imitate, we assume this interaction pattern is simple, fixed, innate and the same for all agents. The interaction pattern was designed such that no external observer is required to judge the success of the game. The imitator decides on the success of the game by comparing its initial action with its best matching action for the observed imitation. Opposed to other approaches, no threshold is required to decide on the success of the game. Due to the categorical perception, the observed action can be compared directly to previously stored actions.

After the interaction, the imitator sends binary feedback to the imitator. Therefore, we assume a single bit perfect communication channel to exist between the agents. Learning consists of two phases: first the imitator adapts its repertoire (step six of the game): If the game succeeds, the action it used is shifted towards the observed action. If the game fails, the same shift is performed on condition that the action considered was not permanently very successful in past interactions. Since it is of no use to adapt a successful action, a new action, matching the observed action, is created in that case.

This adaptation to the imitators’ repertoire is based on the current state of its repertoire, the action observed from the initiator and the feedback it received. This is only local information, i.e. the imitator has no other access to the initiators internal state than by observing its actions and receiving its feedback.

Additionally, both agents perform some general updates on their repertoires (seventh step of the game):

1. With a small probability the agents can add new random actions to their repertoires.

2. With every action, use and success counters are associated. Whenever an action is performed, its use counter is increased. Whenever imitation succeeds, the success counter of the used action is increased. Actions which have proven to be permanently unsuccessful in the past are removed from the repertoire.

3. Actions which are too similar are merged, such that no confusion can exist between those two actions.

Opposed to the update procedures in step six of the game, these three update rules do not depend on the outcome of the actual game, but rather depend on the entire repertoire of the agent.

3. Experiments and results

The imitation game as described above now has at least three different instantiations: it was originally proposed in the study of the emergence of vowel systems (de Boer, 1999). However, in this document, the same framework is used in the context of imitation of actions. Recently, the paradigm was applied to imitation of intentional behaviour (Jansen and Belpaeme, 2005, Jansen, 2005).

In our work on imitation of actions, we use a robot arm with six degrees of freedom. The end-effector of the arm is clearly marked, such that it can easily be observed by a stereo-vision system. Agents perform actions by moving the arm from one configuration to another and thus perform meaningless gestures. When observing an action being performed, the stereo-vision system delivers a sequence of coordinates of the end-point of the robot arm. Using Dynamic Time Warping, observations can be compared. Experiments were performed both using this physical set-up as in a simulated version of this set-up. In this document, all results will be obtained from computer simulations, as that allows us to control the degrees of freedom and the lengths of the joints easily in our experiments on imitation learning among dissimilar embodied agents.

In figure 1 results of a baseline experiment are introduced. All figures in this document will show both the average imitative success and the average number of actions in the repertoires, unless otherwise mentioned. The imitative success is simply the fraction of successful games. In this first experiment, the population is extremely small (only two agents) and the embodiment is very simple as well: both agents have a simu-
lated two-dimensional robot arm with only three joints. The same experimental set-up was used in the work of Alissandrakis (Alissandrakis et al., 2004). This simple experimental set-up will allow us to compare our results with experiments reported later in this document. The figure illustrates the main properties of the game: At the beginning of the game, the number of actions in the repertoires of the agents steadily increases. However, after a while the growing is tempered. The imitative success on the other hand remains very high at any stage of the game. In this document, it is not shown that the repertoires of the agents are actually similar. A measure for sharedness was developed and can be used to illustrate this. For instance in (Jansen, 2003a), this measure is used.

4. Dissimilar embodied agents

![Initiative success and number of actions learnt for two agents playing 5000 imitation games. One agent has two joints, the second one has three joints. The model of the kinematics of the end effector is preprogrammed. Results are averaged over ten runs.](image)

In this section, we investigate whether our framework of imitation games can lead to shared repertoires of actions when agents have different embodiments. Imitation among dissimilar embodied agents is an important problem including several theoretical issues, for instance the correspondence problem (Nehaniv and Dautenhahn, 1998, Dautenhahn and Nehaniv, 2002). However, even when studying imitation between two exact copies of the same robot, issues on dissimilar embodiment are raised since no two robots will have exactly the same sensory-motor behaviour.

Suppose a dolphin makes a waving motion with its entire body. How can this be imitated by a human? By waving an arm (which one?), by moving the entire body, by moving the neck, ... ? In our set-up however, this is not an issue we investigate. We suppose that all robots have very similar characteristics: they all have two-dimensional arms, with a clearly marked end-point. Only the number of joints and the length of the joints can differ. Even in this simplified set-up, a second important issue must be tackled. It is very well possible that actions performed by one agent, can simply not be performed by the other agents. Agents with insufficient degrees of freedom or with joints not long enough, might not be able to reach a certain configuration.

In figure 2, the imitative success and average number of actions are shown for two agents playing imitation games. Both agents have a different embodiment: one of them has two joints, whereas the other one has three joints. The joints of both agents all have the same length. That means that one agent can reach all positions in a circle of radius $2r$, the other agent can reach all positions in a circle of ratio $3r$. Thus, only $\frac{1}{3}$-th of the space reachable by the agent with 3 joints can be reached by the agent with only two joints. However, the imitative success remains very high, showing that—at least under this restricted conditions—imitation among agents with different embodiment is possible without modification to this game. This illustrates our claim that only those actions that can be observed and performed by all agents are actually maintained in the repertoire.

5. Learning a model of the kinematics

According to many developmental theories (both in human development (Rao and Meltzoff, 2003) and robotic development (Jansen et al., 2004, Calderon and Hu, 2005, Manuel Lopes, 2005)), an accurate model of the kinematics of the body of the agent is a prerequisite for successful imitation.

In the imitation game, the imitator requires a model of the inverse kinematics of its own body. When observing an action, he uses the model to deduce the motor commands it must issue itself when performing the same action. A model of the forward kinematics is not strictly required, since an agent can simply perform an action and observe the outcome of the action instead of using the model. So the agent can in fact learn the mapping between actions and states. However, learning this mapping is time consuming and error prone (as it involves perceiving your own state, a process which is inherently noisy). In previous work, we have therefore preprogrammed a model of the kinematics into all agents.

Alternatively, we could take the developmental point of view and first learn the model of the kinematics. In the community of robotics a multitude of techniques were developped, see for instance (Jordan and Rumelhart, 1992). They all use the same paradigm of individual exploration: agents try to reach a certain position in the world, for instance an object. By observing the difference between real outcome and intended outcome, a model can be learnt (this is also
known as motor babbling).

In our experiments, we use locally weighted learning (LWL) (Atkeson and Schaal, 1995), which is an instance based learning method, because it is simple, fast and adaptive. Whenever agents perform actions, they associate performed motor commands and observed actions. Both the motor commands and the observations are kept sorted in every dimension. This allows fast lookup of the nearest neighbours of any motor command or observation using the projection method.

For calculating how a motorcommand \( m \) maps to an observation \( o \), the \( n \) nearest neighbors \( m_1, \ldots, m_n \) of \( m \) are looked up in memory. The observation space is assumed to be linear in the neighbourhood of \( o_1, \ldots, o_n \). This way, a linear mapping from motorcommands \( m_i \) to observations \( o_i \) can be calculated in the neighbourhood of \( m \) using single value decomposition (SVD). The importance of the motorcommands \( m_i \) in the mapping can be weighted according to their distance to the query point \( m \).

As expected, a model of the kinematics could be learnt easily. The agents need relatively few associations for obtaining good results in learning both the forward and inverse kinematics of their body. There is no difference in imitative success with a preprogrammed or a learnt model of the kinematics. Results are reported in (Jansen et al., 2004).

More interesting is the experiment reported here. We investigate whether the agents can learn the model of kinematics while imitating. The agents start imitating without any built in model. They use LWL for calculating forward and inverse kinematics. Initially, this is not possible since the technique requires at least \( n \) associations. In that case a random motorcommand or observation is returned.

While trying to imitate each other's actions, the agents observe themselves performing actions and gradually provide training instances for the LWL. Initially, imitative success is a little bit lower, since intended and actual behaviour can differ as the model of kinematics is still very inaccurate. However, the imitative interactions lead to repertoires of actions that can be observed and performed by all agents. As such, only actions that can easily be distinguished, even when performed inaccurately, are initially learnt.

In figure 3 the imitative success and number of actions for two agents playing 5000 imitation games are shown. The agents started without a model of the kinematics of their body. The model was learned during the imitative interactions. The first agent has only two joints while the second one has three. In figure 4 the number of associations that are stored in the model of the agents are shown. Although imitative success is lower than in the previous experiments, it is still well above random. Results thus suggest that it is possible to start to try to imitate, even if a model of the kinematics is not yet established. Imitation can then be used in stead of individual exploration for learning the model.

6. Evolving agents

In some robotic applications, the kinematics of the body of the agents might unexpectedly change over time (a joint might get stuck) or the body itself might even change drastically (self reconfigurable robots (Murata et al., 2004)). In such cases, the agent must not only be able to learn a model of its kinematics, it must also be able to adapt to its changing embodiment, just like infants do. This can be accomplished in our imitative framework by using an adaptive memory in which older associations are weighted less than newer ones in the LWL, before they are finally removed when not taken into account anymore.

In the experiment shown in figure 5 the embodiment of the agents slightly changes over time, i.e. the agents grow. Agents start with two joints of length ten. However, every 500 games, their joints are increased with one unit. This requires them to adapt their model of their
kinematics in order to remain successful in imitation. Results clearly indicate that the interaction pattern defined by the imitation game is powerful enough to allow for successful imitation in growing agents.

7. Games without feedback

The influence of feedback on the learning of shared categories has been heavily debated in different domains; in (Chouinard and Clark, 2003) it is discussed whether children use only positive feedback or also negative feedback when learning a language. In (Steels and Belpaeme, 2005) it is for instance investigated how a perceptually grounded categorical repertoire can become shared among the members of a population to allow successful communication.

Our paradigm contrasts with classical imitation approaches in which a single teacher and student are considered. In that case the teacher possesses a repertoire of actions, which is transferred to the student by demonstration and observation. In this classical approach, it is obvious that all learning is performed by the student.

In the game as described in section two, most of the learning is done by the imitator, which was motivated by this rationale. However, in our setting in which all agents gradually develop repertoires of actions, the imitator might not be the best suited agent for learning anymore. Moreover, the imitator adapts its repertoire partially based on the outcome of the game, which requires a single bit perfect communication channel to exist between demonstrator and imitator.

We investigate whether variations to the game can be designed in which no feedback is required: in a first experiment it will be investigated whether it is important which of both agents performs the updates which are independent of the outcome of the actual game (step 7 of the game). In the second experiment, we investigate whether games can successfully be played in which it is not the imitator but the initiator who adapts its repertoire based on the outcome of the actual game.

Experiment 1 In the seventh step of the game, both agents perform some general updates to their repertoires: adding new random actions, deleting unsuccessful actions and merging similar actions. We can discriminate between four variants of the game: none of the agents perform these updates, only the initiator performs these updates, only the imitator performs these updates and both agents perform these updates.

Figure 5: Imitative success and number of actions learnt for two agents playing 10000 imitation games. Both agents have two joints. The model of the kinematics of the end effector is learnt using LWL. Every 500 games the length of the joints of the agents is increased with 1 unit. Results are averaged over ten runs.

Figure 6: Imitative success for five agents playing 5000 imitation games. Four variants are investigated, each showing results for different agents performing the updates specified in the seventh step of the imitation game.

Figure 7: The average number of actions in the repertoires of the agents for the four variants shown in figure 6.

In figure 6 and 7, the imitative success and average number of actions in the repertoires of the agents are shown for all four variations mentioned above. Several conclusions can be drawn from this experiment. Obviously, in the first variant, no repertoire of actions emerges. Since agents only have a single action in their repertoires, imitation can never fail and success is maximum. This trivial case will further be ignored. In every of the three other types of games a stable repertoire of actions emerges: the number of actions steadily increases and finally converges to a stable number. Moreover, in
every of the three types, imitation is almost always perfect\textsuperscript{1}, without a significant difference between the success in the different types of the games.

The only difference in performance which can be observed is that the size of the repertoire initially grows faster when both agents perform the updates. Indeed, if both agents perform the updates, the probability of adding a new random action is doubled, explaining the faster increase. After a while, the repertoires of actions contain that many actions that some of them become too similar, causing them to be merged. Therefore, the final size of the repertoire after convergence is similar to the sizes of the repertoires in the other games. From this experiment we conclude that it is not important whether the initiator or the imitator actually performs the updates described in the seventh step of the imitation game.

**Experiment 2** In a second experiment, we assume that both the initiator and imitator perform the updates which are independent of the actual game. Under that assumption, we investigate which of both agents best performs the update described in the sixth step of the imitation game, which is dependent on the outcome of the actual game.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure8.png}
\caption{Imitative success for five agents playing 5000 imitation games. Four variants are investigated, each showing results for different agents performing the updates specified in the sixth step of the imitation game.}
\end{figure}

In this experiment, we discriminate again between four different conditions: none of the agents perform the updates of the fourth type, only the initiator does so, only the imitator does so and both agents do so. Results are shown in figures 8 and 9. They indicate that in all of the four conditions, the agents succeed in developing a stable repertoire of actions. The imitative success is high in all conditions, except in the condition where none of the agents do the updates. However, in the latter case imitative success is still well above random\textsuperscript{2}. So, with learning purely based on score keeping, imitative success is about 70%.

From the combination of the two experiments, several conclusions can be drawn:

1. If at least one of the two agents performs the updates specified in the seventh step of the game, then imitation games with above random success can be played, even if no agent performs updates specified in the sixth step. Our experiments show that purely trial-and-error learning of a repertoire of actions is possible. While the same amount of interactions is required to build a stable repertoire, the imitative success is lower than in the case of learning based on local information consisting of the outcome of the actual game.

2. Better imitative success is obtained if at least one of both agents also performs the updates specified by the sixth step of the imitation game. However, it is not important which of both agents does so, nor is the success higher when both agent do so. A scenario in which the initiator rather than the imitator performs all updates is thus perfectly possible. In that case, all learning is done by the initiator\textsuperscript{3}. This means that the feedback from initiator to imitator about the outcome of the game becomes obsolete.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure9.png}
\caption{Average repertoire size for five agents for the same four conditions as shown in figure 8.}
\end{figure}

\section{Conclusion}

In this paper, we have explained our framework for imitation of actions. It differs from many other frameworks in that it does not assume a fixed teacher-student relation in which knowledge is passed from the former to the

\textsuperscript{1}The success of the experiments in figure 6, 7 and 8 is close to 100\%, while the success in former experiments fluctuates around 80\%. This high success is due to the simpler embodiment of the agents: for the purpose of simplicity and computational efficiency, the agents have only 1 degree of freedom. Since a property of the interaction pattern is investigated, the exact type of actions being imitated is not important.

\textsuperscript{2}In case of imitative interactions with random repertoires, imitative success is about fifty percent. Detailed motivation including proof can be found in (de Boer, 1999).

\textsuperscript{3}Therefore, we have used the terms initiator and imitator, rather than teacher and student.
latter. We rather consider agents which all start without any built in actions. By inventing new random actions and imitating each other, a shared repertoire of actions emerges.

We investigated three assumptions which were implicitly made in our framework. These are: (1) all agents have the same embodiment, (2) all agents have a model of the kinematics of their body and (3) there is a perfect single bit feedback channel between both agents. It is shown that none of those assumptions is crucial for our paradigm of imitation games. This allowed us to show successful imitation among agents with a different embodiment, to show that agents can engage in imitative interactions while learning the model of the kinematics of their body. Finally, we showed that the initiator, rather than the imitator can adapt its repertoire, such that no feedback is required.

Acknowledgments

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

References


