Human and Artificial Agents Learning Categories in Interaction

Giuseppe Morlino[1], Sascha S. Griffiths[2], Lars Schillingmann[2],
Stefano Nolfi[1], Britta Wrede[2], Katharina Rohlfing[2]
[1] Institute of Cognitive Sciences and Technologies – CNR, Via S. Martino della Battaglia 44, 00185 Rome, Italy
[2] Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany
{giuseppe.morlino, stefano.nolfi}@istc.cnr.it, {sgriffit, lschilli, bwrede, rohlfing}@techfak.uni.bielefeld.de

We present a set of experiments on category learning in which a human or artificial agent has to learn to master a task that requires to properly manipulate unknown objects. The agent learns to master the task by exploiting the feedback provided by a tutor. These experiments extend Morlino et al (2010) to take into account how social interaction bootstraps the acquisition of semantic knowledge. They aim at showing how a tutor can actually aid and improve the learning of categories in both artificial and human agents. We hypothesize that the social interaction required for the acquisition of categories needs to be established via contingency (DeJaegher, Di Paolo and Gallgher 2010). We operationalise contingency as synchronous feedback as opposed to asynchronous feedback. We aim to analyse the characteristics of the feedback provided by the tutors and the way in which such feedback is beneficial to the agent's learning process.

In the human experimental scenario, the participants are a human tutor and a human learner. The learners are asked to interact with 16 2D-objects using a mouse device. The objects vary with respect to four perceptual binary features (colour, shape, weight and colour blinking). They are grouped into four categories, formed by four objects each, on the basis of the values of two of the four features. Each object has to be manipulated on the basis of the category it belongs to (i. e. it has to be shaken horizontally or vertically or placed in the right or lower-left position of the screen). The experiment involves a learning agent, located in front of a computer, who should discover how the objects have to be manipulated (he/she does not know the number and the type of categories and behaviours, nor the relation between them) and a tutor, observing the learning agent from a different room, who already knows how the objects have to be manipulated. In a first series of experiments the tutor is allowed to interact with the learning agent by turning on/off 6 symbols (constituted by stars of different colours) on the learning agent's computer screen. The experiment is repeated in a synchronous and asynchronous condition in which the tutor is allowed to select/deselect the symbols also while the learning agent interacts with an object or only before the learning agent starts to interact with an object, respectively. In the artificial experimental scenario the task is the same. Here the tutors are implemented through a set of pre-programmed rules and the learner is an artificial neuronal network trained through a simple adaptive algorithm. In a second series of experiments, the tutor will be allowed to interact with the learning agents sub-symbolically by moving a pointer on the learning agent's screen. Additional explanatory material is available in Morlino et al (2010) and from http://laral.istc.cnr.it/esm/morlino-etal-ICCNS2011.

Foreseen results. We expect that successful tutors might use the possibility to interact "symbolically" in a variety of ways. Some tutors for example might select one symbol during the entire duration of a trial (object manipulation) to indicate the category of the object. Others might select a symbol at a certain point to label a useful behaviour produced by the agent (for example an oscillatory behaviour) independently of whether the behaviour is produced for the right object and at the beginning of a trial to encourage the agent to produce the corresponding appropriate behaviour. Others might use two symbols to express yes/no (good/bad) judgements. In general we expect that the way in which the tutor will use symbols will be influenced by the learning agent's behaviour and by the way in which the learning agent reacts to the selected/deselected symbols. Moreover we expect that synchronous interactions will result in better performance with respect to asynchronous interactions. The analysis of the obtained results and of the type of feedback produced by the tutors should provide hints on how human/artificial agents discover categories and behaviours, on how such learning process can be enhanced by interaction, and on the effect of symbolic/sub-symbolic communication channel. Moreover, we expect that the comparison between experiments performed with human and artificial agents/tutors can help to model human learning and to design more effective artificial learning agents.

References

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in cognitive sciences*, *14*(10), 441-447. doi: 10.1016/j.tics.2010.06.009.

Morlino, G., Gianelli, C., Borghi, A. M., & Nolfi, S. (2010). Developing the Ability to Manipulate Objects: A Comparative Study with Human and Artificial Agents. *Processings of the Tenth International Conference on Epigenetic Robotics* (pp. 169-170). Örenäs Slott, Sweden.