2014 Joint IEEE International Conferences on Development and Learning
and Epigenetic Robotics (ICDL-Epirob)
October 13-16, 2014. Palazzo Ducale, Genoa, Italy

Testing Piaget's Ideas on Robots: Assimilation and Accommodation Using the Semantics of Actions

Eren Erdal Aksoy, Markus Schoeler and Florentin Wörgötter
Georg-August-Universität Göttingen, BCCN, Department for Computational Neuroscience
Inst. Physics-3, Friedrich-Hund Platz 1, D-37077 Göttingen, Germany
Email: [eaksoy,mschoeler,worgott]@gwdg.de

I. Introduction

Defining a generic action representation to learn the variations in trajectory, pose, and object phases is of great interest in cognitive robotics. Conventional methods are insufficient or slow for efficient human-like action understanding and generation in robots. This is due to the fact that the same action (e.g.*"cutting"*) can be performed with a wide variety of tools (*knife or cleaver*), on many different objects (*meat or rubber)* and in various ways, e.g. using different movement patterns. In spite of this high-dimensional complexity, children can understand and learn the *"meaning"* of an action, whereas robots so far cannot.

We have recently introduced Semantic Event Chains (SECs) [1] as a novel concept to encode the semantics (*"essence"*) of manipulation actions independent from all variations in trajectory, pose, and object domains. SECs are derived on-the-fly from the graph representations of the consistently tracked object segments in the scene. Graph nodes hold segment centers and edges indicate whether two objects touch each other or not. SECs essentially extract the sequence of changes of the spatial (*i.e. touching*) relations between manipulated object segments and are invariant to the particular objects used, the precise object poses observed, the actual trajectories followed. In [2] we showed how to enrich event chains by incrementally appending observed trajectory, pose, and objects like action descriptors only at decisive instants that are naturally emerging in the columns of SECs. Hence, the cognitive agent can employ the enriched SEC representation to efficiently imitate complicated actions even under different circumstances as shown by various experiments in [3].

II. Method

We here aim at using the framework of enriched SECs to implement two learning concepts from child psychology (Accommodation and Assimilation suggested by Piaget [4]) in artificial agents to equip them with the means to acquire action understanding. Assimilation is a refinement process for memorizing action representation (schema) created for a learned action. On the other hand, accommodation happens when the agent realizes that the new observed action does not match any of its known schemas and hence requires a novel schema to be learned. SECs are here considered as action schemas and the semantic similarity, defined in [1], between each is employed to decide whether to create a new memory unit (Accomodation) or to update the most similar memorized schema (Assimilation) with the novel syntactic details, i.e. object and trajectory information.

We benchmarked the proposed framework with a large action dataset, introduced in [5], consisting of eight manipulations: *cutting, chopping, stirring, pushing, hiding, putting, taking, and uncovering*. Each manipulation consists of 15 versions demonstrated by 5 different individuals. The dataset has in total 30 various objects manipulated in all these 120 demonstrations. Fig. 1 shows the corresponding schemas (SECs), created for *cutting* and *stirring* actions, together with the estimated syntactic details, i.e. object and trajectory features. We employed the model-free incremental learning framework, introduced in [5], to create a SEC model for each observed manipulation type without requiring any human intervention. Once a novel version of a previously learned action model was detected by means of the semantic similarities, we applied the instance based object recognition method from [6] to image segments, i.e. graph nodes in SECs. In this way, we link actions with objects manipulated in all different demonstrations of the same action type. We further decomposed the trajectory information of the hand into shorter fragments by considering the decisive temporal points represented by the SEC columns. Instead of storing the complete trajectory, the artificial agent, now, captures the crucial trajectory interval while the object is being manipulated, e.g. the cutting interval while *the knife is touching the cucumber* or the stirring instant while *the spoon is touching the milk* as highlighted by gray trajectory boxes in Fig. 1. Extracted trajectory segments were encoded by the modified Dynamic Movement Primitives (DMPs, [7]), coefficients of which were compared with four predefined movement types: *linear, cubic, parabolic, and elliptic*.

Among eight manipulation types in the dataset, the proposed learning algorithm extracted seven SEC models by naturally merging the *cutting* and *chopping* manipulations (assimilation!) due to high semantic similarity between each task. Fig. 2 displays the classification accuracy (on the left) of 120 demonstrations compared to those learned SEC models. Observing no misinterpretation between tested and learned manipulations proves the strength of the proposed semantic representation. Fig. 2 also depicts the usage rates of the primarily manipulated objects (in the middle) and the mainly followed trajectory types (on the right). For instance, in the *stirring* action not only *spoon* but *knife* were also used mainly with *parabolic* trajectory types. However, in the *cutting & chopping* task *knife*, *spatula*, and *cleaver* were chosen with the *parabolic* and *cubic* trajectories.

978-1-4799-7540-2/14/$31.00 ©2014 IEEE

III. Conclusion

The proposed framework addresses the problem of implementing a high level *"psychological"* learning mechanism at the level of machines. The learning framework is bootstrapped with the semantic relations (SECs) between observed manipulations without using any prior knowledge about actions or objects while being fully grounded at the sensory level (image segments). With the concept of employing the semantics of actions and linking with the syntactic action descriptive information, a growing (cumulative) memory of actions can be efficiently developed in a machine similar to processes suggested for human learning by the psychologist Jean Piaget in 1953 [4].

Acknowledgment

This work was supported by the EU Cognitive Systems project Xperience (FP7-ICT-270273).

References

[1] E. E. Aksoy, A. Abramov, J. Dörr, K. Ning, B. Dellen, and F. Wörgötter, "Learning the semantics of object-action relations by observation," *The International Journal of Robotics Research*, vol. 30, no. 10, pp. 1229–1249, 2011.

[2] E. E. Aksoy, M. Tamosiunaite, R. Vuga, A. Ude, C. Geib, M. Steedman, and F. Wörgötter, "Structural bootstrapping at the sensorimotor level for the fast acquisition of action knowledge for cognitive robots," in *IEEE Int. Conf. on Development and Learning and Epigenetic Robotics (ICDL-EPIROB)*, 2013.

[3] M. J. Aein, E. E. Aksoy, M. Tamosiunaite, J. Papon, A. Ude, and F. Wörgötter, "Toward a library of manipulation actions based on semantic object-action relations," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013.

[4] J. Piaget, *The Origins of Intelligence in the Child*. London, New York: Routledge, 1953.

[5] E. E. Aksoy, M. Tamosiunaite, and F. Wörgötter, "Model-free incremental learning of the semantics of manipulation actions," *Robotics and Autonomous Systems (RAS) (under review)*, 2014.

[6] M. Schoeler, S. Stein, J. Papon, A. Abramov, and F. Wörgötter, "Fast self-supervised on-line training for object recognition specifically for robotic applications," in *International Conference on Computer Vision Theory and Applications VISAPP*, January 2014.

[7] T. Kulvicius, K. J. Ning, M. Tamosiunaite, and F. Wörgötter, "Joining movement sequences: Modified dynamic movement primitives for robotics applications exemplified on handwriting," *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 145–157, 2012.

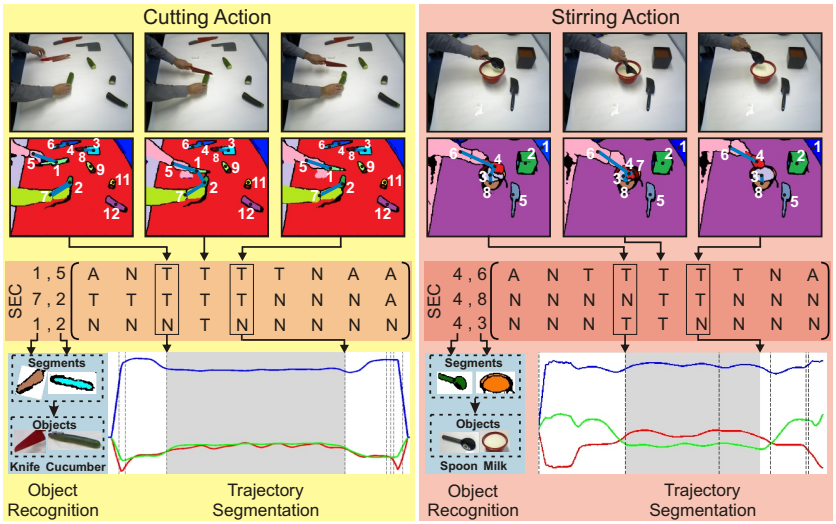

Fig. 1. Enriched SEC representations for two real action scenarios: *"Cutting a cucumber with a knife"* (on the left) and *"Stirring milk with a spoon"* (on the right). Each column corresponds to one key frame, some of which are shown on the top with original images, respective segments (colored regions), and graphs. Rows are spatial relations between object pairs, e. g. between the knife (1) and hand (5) in the first row of the SEC for the *cutting* action. Possible spatial relations are N, T, and A standing for *"Not-touching"*, *"Touching"*, and *"Absence"*. On the bottom, object identities (derived from image segments) and decomposed trajectory fragments are displayed. Boxes in the SEC columns show temporal borders of the trajectory segments indicated in gray.

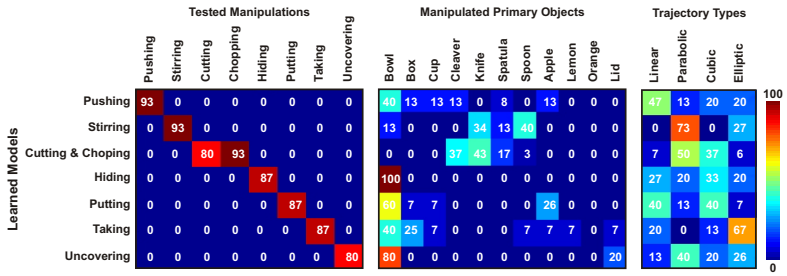

Fig. 2. Confusion matrices showing both the classification accuracy (on the left) for the complete data set including in total 120 manipulation samples, the usage rate of different primarily manipulated objects (in the middle) and and the abstract trajectory types (on the right) chosen in each learned action model.