Learning Simultaneous Sensory and Motor Representations- (SESEMO)

Berkeley Kids

Redundancy reduction [1], edge detection [3] and hierarchical representations [5] have been the main stay for lot of work in computer vision and vision neuroscience to represent sensory data. But, these representations do not directly lend themselves to action. It has been hypothesized [6] that brain probably represents sensory information in a way that helps an organism to act in the world.

Philipona and O'Regan [7, 8] showed that the joint manifold of sensory and action space taken together has a lower dimensionality than that obtained by simply adding the dimensions of motor and sensory spaces. An observation which explains this is that motion of an object in the world and the equal but opposite motion of the organism with the object being stationary leads to the same sensory percept. It's the relative motion which matters. In other words, sensory and motor spaces are 'compensable'. This insight is probably employed by organisms to achieve stability of percept. For example, even when we humans are moving we perceive the world to be stationary.

In this work we consider the problem of joint learning of sensory and motor representations within the context of percept stabilization. This goal is interesting because it lets the agent to build a sensory representation that is coupled with its actions. We assume that agent has no apriori knowledge of its kinematic model nor does it have any meaningful representations of its sensory stimulus, i.e. the agent is in the state of 'tabula rasa' initially.

In overactuated motor systems (high degrees of freedom/redundancy) training a control model becomes challenging. [9] explored learning of inverse kinematics to reach a certain goal state. Other works have proposed solutions based on the concepts of motor primitives [10]. However, no work has looked at learning such primitives from the statistics of the motion that an agent needs to perform to achieve a goal. Here, we wish to explore the idea of learning motor primitives that can then be composed sequentially to perform actions. Such a representation regularizes the space of possible motor actions and allows the agent to conduct high level planning of actions by providing it with a vocabulary of primitive motions to chose from. That is to say, at each time step we no longer have to specify the control sequence for each controller but the problem is now of choosing an appropriate motion primitive (aka basis element). This is also a biologically plausible system, especially considering the fact that spinal reflexes (akin to motion primitives) and cortical motor systems (high level planning) work closely to make complex motions possible[4].
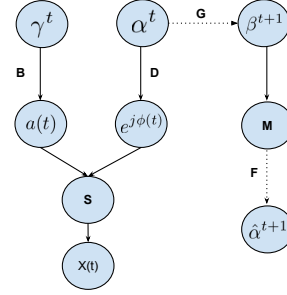

Fig. 1: This figure depicts a graphical model representation of the sensory-motor system of the agent. Please refer to the text for more details. Note, the term graphical model is used in a loose sense.

It seems likely that for a robotic agent, a hierarchical control system with some low-level task dependent reflex circuits that are modulated by high level commands will lead to easier planning. Work of [11] in optimal control theory literature also talks of related ideas.

The main ideas proposed in this work can be summarized as following:

- Learning of a motor basis for control (just like gabor functions form a basis for image patches.)
- Goal dependent joint estimation of sensory and motor representations.
- Developing adaptive fault-tolerant sensorimotor representations.

I. Model

Fig 1 depicts the sensory-motor system of the agent. $\mathbf{X(t)}$ is the sequence of images (frames) that fall on the retina (say). We then learn a sparse generative sensory representation along the lines of Cadieu & Olshausen [2] that tries to account for both form (γ^t) and motion(α^t) separately using complex basis elements. We refer the reader to [2] for more details for its training.

$\mathcal{G}$ transforms the motion component of the sensory percept to the motor primitive space β^{t+1} and the input to the actuators is given by $M\beta^{t+1}$.

$$\beta^{t+1} = \mathcal{G}\alpha^t$$

In order to stabilize the percept, at every time step the agent must move in a way which would compensate for the change

in visual stimulus caused due to motion of objects in the environment. Consequently, at every time step, the agent tries to predict the sensory percept which it expects to see at the next time step(i.e. $\hat{\alpha}^{t+1}$). Note, in contrast to predicting the next image frame in the pixel space, the agent predicts in the parameter space - i.e the agent predicts the sensory percept it would infer at the next time step instead of the pixel values themselves. This prediction is expressed as following:

$$\hat{\alpha}^{t+1} = \sum_{i=0t}^{t-1} H_{t-i}\alpha(t-i) + \mathcal{F}\mathcal{M}\beta^{t+1}$$

where, $\mathcal{F}$ is the transformation from the motor space to sensory parameter space and H_{t-i} is the effect of the sensory parameters inferred at time t-i on the prediction of time point t+1.

$$\min_{F,M,G,\tau} \|\alpha^{t+1} - \hat{\alpha^{t+1}}\|_2 s.t. \tag{1}$$

$$\sum_i \lambda_2^i G^i \tag{2}$$

where β^{t+1} is computed as follows

$$\tag{3}$$

The objective that we wish to minimize is defined in Eq 1. Eq 2 defines the constraint that we require the columns of G (defined by G^i to be sparse so that our inferred α results in a sparse β. At this point, we explore no explicit constraints on F other than those specified above. We compute β^{t+1} as a transformation of α^t as described in Eq 3. We compute our expected value of $\hat{\alpha}^{t+1}$ as a weighted sum of previous α^{t-i} and a transformation of our current action into a sensory percept by the matrix $\mathcal{F}$ as seen in Eq 4 where the exponent helps to impose a decaying weight for oldest time frame.

II. Discussion

III. Acknowledgments

All the code for this project can be found on our github[1] link.

References

[1] Horace B Barlow. Possible principles underlying the transformation of sensory messages. *Sensory communication*, pages 217–234, 1961.

[2] Charles F Cadieu and Bruno A Olshausen. Learning intermediate-level representations of form and motion from natural movies. *Neural computation*, 24(4):827–866, 2012.

[3] David H Hubel and Torsten N Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1):215–243, 1968.

[4] Eric R Kandel, James H Schwartz, Thomas M Jessell, et al. *Principles of neural science*, volume 4. McGraw-Hill New York, 2000.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, volume 1, page 4, 2012.

[6] J Kevin O'Regan and Alva Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24(05):939–973, 2001.

[7] David Philipona, J Kevin O'Regan, and J-P Nadal. Is there something out there? inferring space from sensorimotor dependencies. *Neural computation*, 15(9):2029–2049, 2003.

[8] David Philipona, Jk O'regan, J-P Nadal, and Olivier Coenen. Perception of the structure of the physical world using unknown multimodal sensors and effectors. In *Advances in neural information processing systems*, page None, 2003.

[9] Matthias Rolf and Jochen J Steil. Efficient exploratory learning of inverse kinematics on a bionic elephant trunk. 2012.

[10] Stefan Schaal, Jan Peters, Jun Nakanishi, and Auke Ijspeert. Learning movement primitives. In *Robotics Research*, pages 561–572. Springer, 2005.

[11] Emanuel Todorov. Optimality principles in sensorimotor control. *Nature neuroscience*, 7(9):907–915, 2004.

[1] github.com/rctn/sesemo