Indian Institute of Technology Kanpur

INTRODUCTION TO COGNITIVE SCIENCE

SE367A

Visuomotor Learning Using Image Manifolds: ST-GK Problem

Author: ANURAG MISRA Department of Computer Science and Engineering Supervisor: DR. AMITABHA MUKHERJEE Department of Computer Science and Engineering



Abstract

Vision provides us with much information about distal spaces, and enables us to formulate expectations about how objects move. In this work, we consider how images, together with motor and performance data, can be used to form "maps" - or a holistic view of any motion. For example if the learner knows how to throw a ball, it may be able to use the map to see that it also applies to kicking, or catching, or throwing darts. These maps, we suggest, are manifolds that constitute dense latent spaces for these motions. While earlier approaches to learning such systems (e.g. discovering laws of physics) used prior knowledge of the set of system variables, we assume no such knowledge and discover an equivalent representation with alternate parameters (the latent parameters of the non-linear manifold). We show how such a system can be built without any explicit parametrization for a number of motion systems from rolling on an incline to pulleys to projectile motion. We then demonstrate how such a map (for projectiles) can be used for solving the standard problem of a striker kicking the ball against a goalkeeper towards the goal, and how practice can improve the relevant part of the map, and hence the throwing capability.

1 Introduction

Robots lack the power to learn things from the observations and experience, similar to how babies learn. Consider a baby robot that observes how things move in reaction to its actions or those of others. Smith and Gasser state that the learning procedure of babies start from observing events in their environment and gathering information. The sensory system collects all the required information by moving body parts and recording the experience in terms of perception or sensory input. Of all the sense contributing to gather information to learn about a system, vision is our most important sense. It appears to have a key role in forming the spatial knowledge in the brain. In our work, we are trying to learn such a basic map behind various systems using visual input. We build a generalised system that is able to understand the 3 overall functioning of any physical system or motor action observed. All the skills are correlated to each other in some way or the other. We have an overall general map defined for them in the brain such that a person possessing one particular skill will be able to perform an action involving the use of the other related skill. For example, a baby knowing how to throw is also able to catch, kick or shoot with some precision. We are trying to learn that general map formulated in our brain. Figure below shows how the two actions of throwing and kicking can be mapped to the same visual input.



(a) Action of Throwing ball

(b) Action of kicking ball

The patterns in both the actions is the same and we are trying to discover such patterns in our model which can then be finetuned accordingly to expertise gained in one particular skill.

Indeed, primate brains operates in a very integrated manner. Motor behaviour invokes visual simulation to enable rapid detection of anomalies. while visual recognition invokes motor responses to check if they match. Each time the robot throws the ball for motor learning, it also generates a visual trajectory from which it can learn a visual model of the action. Further, the motion parameters are correlated with the visual feedback - both lie on matched low-dimensional curved manifolds which can be aligned. We develop a unified approach for modeling actions, based on visual inputs (say trajectories or paths) resulting from given motor commands. The model constitutes a low-dimensional manifold that discovers the correlations between visual and motor inputs. The model can be applied to either visual recognition or to motor tasks. The attempt of this work is to build a model similar to the model of human brain for the skills combining vision and motor actions. The model built is then used as starting point to learn specific tasks for example, the projectile motion system can be used to learn the aiming tasks like striker goalkeeper problem.

1.1 Past Work

Past works in this field are related to this work in mainly two ways, one is learning physical laws from observations and the other is optimising model to attain motor skills like throwing, catching, kicking etc.

1.1.1 Learning Physics Laws

Past works in this field learn from the experimental values either specified as input or recorded by the system itself. These give the idea of what methods can be used to generalise the variability noted in a system in terms of data values. Our work also aims at capturing that variability but the observations are in terms of perception. This includes inputs in terms of images or videos or even visual sensory data that is being gathered by the person when he/she observes a system in working or performs it itself. One more difference is that our system does not make use of any priors about the system. Our model is a generalised one which can be applied to any system and information about that system can be derived.

1.1.2 Learning Motor Behavior

For every motor action we perform, our brain formulates an internal model and then predicts the outcome of that model. The predicted outcome is then matched with the feedback calculated from the perception of the action. The error is then used to modify the internal model for the same. It has also been shown in how action and perception are closely related. This work shows that even when one observes a motor action taking place, the brain triggers the same neurons that are triggered when one performs that action himself. This is the reason why one can predict the outcome when shown with a video clip of a player kicking the ball towards the goal. The accuracies were high for the videos of themselves because their brain was able to match the perception with the exact action it performed.

In humans, the models for the motor actions start developing in the brain from 8 a young age. The babies slowly learn to perform actions like walking, throwing etc. by building models using observations and personal experience. These models keep on improving with exposure to more data. The model of throwing can be modulated to learn finer skills like throwing a baseball, a frisbee or even a dart, catching, kicking etc. In baseball, making the decision on the fly about the direction to move in order to catch the baseball seems like very difficult to learn. But, it has been proved that the baseball player follow a path to maintain the linear optical trajectory(LOT) of the ball. The work states that the players run in such a manner that the rate of change of the optical angle of the ball remains constant. Dogs also follow the same technique to catch the frisbee thrown at them. This has been shown This current approach is different from all the past methods as neither are we trying to learn the scientific equations of the laws underneath the system nor learning a specific skill or even a specific skeleto-muscular system. // Instead we attempt to learn the generalised map of the actions which can be used for all kinds of activities that involve a visual input and a reaction to that input using any motor action. We are attempting to find out the basic properties that if known about the system can help us control the system in the most optimised way. In our work, the observations are the perception of various physical systems in the form of images. These images are clustered on a lower dimensional manifold to discover the variations and patterns in the system. Similar to most of the past works, we also build a model to describe the patterns but our model is built on the lower dimensional embedding space(manifold) rather than the image space. The analysis of the manifold learnt reveals many important properties of the system. The motor skill of throwing also is associated with the physical system of projectile motion and hence is learnt through the same model. We describe a generalised algorithm that can be applied to the images of any physical phenomenon and will discover the understanding of the underlying law without any prior domain specific knowledge about the system.

1.2 Manifolds and Dimensionality Reduction

Learning a few visuo-motor tasks are among our agent's very first achievements. Let us consider the act of throwing a ball. Our learner knows the motor parameters of the throw as it is being thrown - here we focus not on the sequence of motor torques, but just the angle and velocity at the point of release. We are given a large set of trajectory images (say, N=1080), each with 100×100 pixels. Each image can be thought of as a point in a 10^4 dimensional space. The set of possible images is enormous, but we note that if we assign pixels randomly, the probability that the resulting image will be a throw trajectory is practically zero. Thus, the subspace of throw images is very small. Next we would like to ask what types of changes can we make to an image while keeping it within this subspace? In fact, since each throw varies only on the parameters (θ, v) , there are only two ways in which we can modify the images while remaining locally within the subspace of throw images. This is the dimensionality of the local tangent space at any point, and by stitching up these tangent spaces we can model the entire subspace as a non-linear manifold of the same intrinsic dimensionality. The structure of this image manifold exactly mimics the structure of the motor parameters (the motor manifold). They can be mapped to a single joint manifold, which can be discovered using standard non-linear dimensionality reduction algorithms such as ISOMAP. ISOMAP makes use of classical MDS(Multidimensional Scaling) to find the low dimensional embedding of the data. It tries to preserve the geodesic distance between every pair of points. The steps include constructing a neighbourhood graph based on the distance metric used and computing shortest path between every pair of points. The geodesic matrix is generated by combining the weights of the edges on the shortest path in the neighbourhood graph constructed. The lower dimensional coordinates are found by applying MDS on the geodesic matrix. The geodesics are computed not in the euclidean space but in the curved image space of the data which helps in preserving the geometry of the high dimensional data. This is the reason why it performs well even with highly non-linear data. The computations performed in ISOMAP are not very complex because all the calculations are performed on the neighbourhood graph constructed which is often small, depending upon the value of K. The major percentage of the computation time of ISOMAP is spent in computing the pairwise distance matrix used to find the k-neighbours of every point. The 11 other major contribution is by the Dijkstra's algorithm $(m+n\log(n))$ for m edges and n data points))to compute the shortest distance between every pair of points. For our problems however, the time is low since the number of data points are not very large. Due to the high dimension of the images used in our systems (of the order 10^4), discovering the hidden manifold seems difficult but ISOMAP discovers the intrinsic dimensionality easily and accurately. Our data is densely sampled and is also free from any discrepancies, therefore, ISOMAP performs well with the images in our problems.

2 Mechanics Law Discovery Algorithm

2.1 Problem Statement

We want to find out the basic properties which, if known about the system can help us control the system in the most optimised way. We try to learn the number of degrees of freedom in a system by learning a manifold from the set of images. We learn the controlling parameters of the system solely from the image data. The goal is to gain the implicit knowledge behind the physical system. This knowledge should be such that we are able to perform the task optimally and accurately without knowing the exact equations of the laws underlying the system. The purpose of this thesis is to show how the images of a physical phenomenon can be used to discover latent characterisations leading to laws of the system that mirror discoveries in physics. These images can be clustered on a manifold to discover the controlling parameters of that system as well as the relationships between them. The embedding points are not directly mappable to the control parameters for the system, so, we train a back-propagation network to compute that mapping. The network is trained on the points on which the manifold is built. It can then be used to determine the values of the parameters for other points in the manifold.

2.2 Algorithm and Approach

The process of getting the embedding and training network is outlined in Algorithm 1. The main steps involved in building the model are computing the low dimensional embedding and training the neural networks according to the obtained embedding.

In our work, we are using ISOMAP technique which attempts to preserve the geo-desics such that the points close to each other in the image space are close in the low dimensional space and the points far apart are distant to each other

Algorithm 1 Mechanics Law Discovery Algorithm

- Input: Set of high dimensional images {I₁, I₂, I₃, ...I_N} and corresponding control parameters
- 2: Output: Value of the control parameters for the set of query images
- Step1: Obtain a low dimensional embedding of the high dimensional image data using ISOMAP dimensionality reduction technique.
- 4: **Step2:** Train a regression model to acquire a mapping from the low dimensional (curved)manifold to the respective control parameters.
- 5: Step3:For executing a new throw, use a (query) image with desired path. Find a linear interpolation J for this query image J_i = ∑^k_{j=1} w_jI_j
- 6: Step4: Calculate the embedding points for the query image using the weights learnt in Step3. $\hat{Q}_i = \sum_{j=1}^k w_j \hat{q}_j$
- 7: Step5: Use the mapping learnt in Step2 to obtain the corresponding parameters
- for the query image J_i .
- Step6: Analyse the manifold to depict interesting relations about the physical system.

[†] Using

linear interpolation on the nearest neighbours, the weights are computed for the query images. Then these weights are used to get the embedding coordinates which can then be fed into the neural network trained to obtain the values of the parameters for the query image.

2.2.1 Solving the Striker-Goalkeeper problem using learnt system

Now we consider a refinement of the above neural network to solve the specific problem of striker kicking the ball into the goal againt the goalkeeper.

- Assumptions
 - Goalkeeper still
 - Point of strike fixed
 - GK position: in front of the goal at some distance
- Approach to solve the problem
 - Assume the 2D goal as a grid as given



- Each point as shown in this grid can be considered as pixel of the $(N+1)^*(N+1)$ pixels image/goal
- It is possible for the striker to striker at many of these points considering the position of GK
- Assume there is no goal keeper. For this condition, given the striker position we find the possibly trajectories for each of these points, progressing in a vertical fashion
- For each vertical line in the grid, given the strike position the value of Range R is fixed, which we supply as a parameter
- This approach reduces the problem to 1D, a 1D containing N+1 eligible points for impact. This reduces to Dart Throwing problem:
- Assumptions
 - Dart board fixed
 - Shoot at fixed point on the dart
 - Point of shoot fixed
- Approach to solve the problem
 - Dart Throwing Algorithm

Algorithm 2 Dart Throwing Algorithm

- 1: Input: Image containing the dart.
- Output: Series of trajectories I₁, I₂.. I_p that will hit the dart in the most optimal position.
- Step1: Calculate the errors e for all sets of trajectories and select the ones with error < ε.

- Step4: Plot the embedding points on the manifold to find the region of accurate trajectories (Figure 3.5).
- Step5: Use the mapping learnt in Chapter 2 to generate more throws in the accurate region. Re-calculate errors.
- Step6: Analyse the accurate region on the manifold to select the trajectories with motor parameters that optimise one's effort in throwing.
- After getting all the trajectories for points in the goal, those trajectories which pass through the goalkeeper reach area are eliminated
- This is achieved by obtaining ypos in trajectory for xpos of GK Reach area in the current vertical trajectory plane and checking if that ypos is true for GK Reach Area











4 Conclusion

- A generalized model to discover the visuo-motor patterns from the set of images of physical systems or systems involving motor activities like throwing developed and executed
- The model built requires no prior knowledge and hence can be used to learn the patterns for any system where images capture the inherent variability of the system.
- Using the model we solved our goalkeeper-striker problem which can be made an integral of machine football games where skill improves by training

5 References

- From visuo-motor to language, Deepali Semwal, Sunakshi Gupta, Amitabha Mukerjee, Department of Computer Science and Engineering, IIT Kanpur
- Resources and code available at http://cocosci.mit.edu/
- Resources available at Google
- A Global Geometric Framework for Nonlinear Dimensionality Reduction Joshua B. Tenenbaum, Vin de Silva, John C. Langford