Reinforcement Learning for Active Appearance Model Dataset Selection

André Antonitsch
School of Technology
Pontifícia Universidade Católica do Rio Grande do Sul
Porto Alegre, RS, Brazil
andre.antonitsch@acad.pucrs.br

Abstract

Active Appearance Models have been used widely for facial analysis and medical image segmentation. The statistical model is based on an input dataset, and describes images based on a set of images used for training it. Training time for the model is based on dataset size and redundancy of image features in the dataset adds little information to the model. In this work we experiment with using a Linear Approximation Q-Learning model as a way to select a smaller subset of images and verify if the generated reduced datasets can be refined by a linear approximation learning model.

Introduction

Active Appearance Models (AAM) were created by Cootes et al. (Cootes, Edwards, and Taylor 1998; 2001), originally to fit a synthetic face onto a target face. This statistical model was shown to be useful in several tasks, such as: modeling faces, facial detection, object detection and image segmentation. While AAMs are powerful models for facial modeling and detection, their expressiveness, that is its ability to correctly fit a more diverse set of faces, are extremely tied to the variance contained in the dataset used to train it. This lack of generality comes from the fact that the model assumes the training dataset to be a contained in a space of what is a valid representation of the object it models (Cootes et al. 1995). The training times for AAMs are directly connected with the dataset size used to train it, further benefiting from a reduced dataset for training.

Selecting a subset of images to use from a given dataset is a similar problem to choosing which columns are more relevant to index in a database, which subset of a given set presents the best results when chosen for a given operation. Using reinforcement learning for database tuning was shown by Basu et al (Basu et al. 2016).

In this project we propose using Q-Learning for dataset selection and pruning for use with AAMs. To accomplish this task, we propose a linearly interpolated approximation as shown by Tsitsiklis et al.(Tsitsiklis and Van Roy 1997). And evaluate whether these techniques are capable of reducing training times for AAMs while providing increased expressiveness with a reduced dataset.

To evaluate the subsets produced by our learning model, we trained several AAMs and tested those against a testing set of facial images. The proposed model was not successful in obtaining a better subset of images as the training progressed and state space was explored.

The work is organized as follows: Section describes the techniques used in this paper and the dataset we use. Section formalizes our problem and our proposed model, the state description, action description, the reward and feature functions we utilized, and the algorithm we used to train our model. Section describes the experiments we performed and results we obtained. Finally, section presents our conclusions, planned future works and final remarks.

Background

This section describes both techniques we plan on exploring, Active Appearance Models and Reinforcement Learning, as well as giving details on how we plan to evaluate the produced model.

Active Appearance Models

Interpretation-through-synthesis is an approach for analysis where the recreation of an object, or subject, through the model can lead to higher level interpretations of the features of said object. The Active Appearance Model proposed by Cootes et al. (Cootes, Edwards, and Taylor 1998; 2001) is one of such models. The AAM can be used for image analysis through the reconstruction of subjects in images, the original work presented was focused in the reconstruction and detection of faces, but its uses extend further, such as image segmentation in medical images and facial recognition.

The AAM combines a model of shape variation and texture variation in images annotated with relevant corresponding landmarks, this work focuses on the shape variation aspect of the model. For example, to build a model for faces, as we propose on this project, one might annotate the outlines of eyes, mouth, face and nose, the method assumes certain landmarks move together as the object moves (Cootes, Edwards, and Taylor 1998; 2001). The annotated points of each

image are aligned in a common co-ordinate frame and each can be described by a $2n$ dimensional shape vector x, where n is the number of landmarks in the image. We then compute the deviation to the mean for each shape vector x, and build a $2n \times 2n$ co-variance matrix of the shape deviation vectors. By applying PCA to the co-variance matrix, we can approximate any example in the training set to the equation 1:

$$x = \bar{x} + P_s b_s, \tag{1}$$

where $\bar{x}$ is the mean shape vector in the dataset, P_s is the set of orthogonal modes of variation obtained and b_s is a set of shape parameters.

The model assumes the landmark points contained in the dataset define an $2n$ dimensional *Allowable Shape Domain* (Cootes et al. 1995). This space describes a region in the where each point is assumed to be a valid and possible shape. This space is assumed to be ellipsoid.

The model expressiveness is the model's ability to correctly fit and describe a more diverse set of images. The expressiveness is then directly tied to how varied the dataset is, given more dissimilar points cover a larger *Allowed Shape Domain*, and therefore represent a larger possible space of valid examples. The model's computational cost also increases with the quantity of training images, therefore the use of redundant images both increases its training time, and does not add to the expressiveness of the model.

For this project we used the Menpo Project (Alabort-i-Medina et al. 2014) open source implementation of an AAM.

Reinforcement Learning

Reinforcement learning is an approach to how agents can learn optimal policies for action taking in a given environment (Sutton and Barto 2018; Kaelbling, Littman, and Moore 1996). Generally, reinforcement learning problems are modeled as Markov Decision Processes (MDP), where a given agent tries maximizing ~~a~~ their accumulated reward given a certain reward function. For this project we plan using the Q-Learning algorithm for policy learning.

Q-Learning maps an action for each possible state of the state space, this is done by iteratively applying an update function to the current mapping, whenever the agent makes an action and observes a certain result. The algorithm then maps a certain reward for taking an action a in a given environment state s as a value pair state-action, usually in a look-up table. The main advantage of this algorithm is the possibility of use in an environment where the reward function is unknown, but measurable in someway.

Ideally, Q-Learning will map every possible state-action value pair into a reward value. However, for environments which are too large or too costly to map exhaustively, alternatives have been proposed. Instead of using a look-up table, state-action value pairs can be a linear interpolation of a given feature set of a given state(Tsitsiklis and Van Roy 1997). The linearly approximated Q-Learning will the approximate the estimated reward value to a linear function as seen in equation 2 (Sutton and Barto 2018):

$$Q(s,a) \leftarrow w^T \cdot F(s,a) + w_b, \tag{2}$$

where w_T is a weight vector, $F(s,a)$ is a function which maps a state and action into a given feature set and w_b is a bias weight.

The Q-Learning update step instead of updating the value associated with the value pair in the look-up table, updates the weight given to each feature in the state's feature set, according to equation 3 (Sutton and Barto 2018):

$$w_i = w_i + \alpha[R(s) + \gamma max_{a'} Q(s',a') - Q(s,a)]\frac{\delta Q(s,a)}{\delta w_i} \tag{3}$$

Selected Dataset

We started by selecting a target database to prune. We selected UTKFace (Zhang and Qi 2017) given its broad range of variation in gender, age, ethnicity and illumination contained in the images. For this experiment, we randomly chose a training set of 11700 images and a testing set of 60 images, both sets contained only images of people between 20 and 40 years old.

Problem Formalization

This section describes our proposed formalization for the task of dataset selection for AAM training, the description for states and actions, the reward function and the training algorithm we used. Given a target dataset of images, we seek to select a subset which offers a good trade-off of training time and expressiveness of the AAM. For these experiments we aimed to sample a set of 500 images from the training set of 11700 images.

States and Actions

Given a set of 11700 training images, and a subset to be sampled of 500 images, there exists $\binom{11700}{500}$ possible combinations, that is, a certainly unmappable quantity of states. To help remedy the exponentially growing combination of both states and actions, we reduced the granularity of our model. To do this, we consider the larger dataset composed of smaller groups of 10 images, and then map our problem to a large set of 1170 groups of 10 images, from which we seek to sample a subset of 50 groups. This reduces our state space to $\binom{1170}{50}$ possible combinations of image groups, which is still very large.

We formalize a state s as tuple (s^+, s^-) where: s^+ is the set containing all the group indexes in the subset selection represented by s, s^- is the set containing all the group indexes of not contained in the subset selection represented by s.

We formalize an action a available in state s as a tuple (x, y), where $x \in s^+$ and $y \in s^-$ and induce a state s' where $x \in s'^- \wedge x \notin s'^+$ and $y \in s'^+ \wedge y \notin s'^-$. The available actions A in a given state s are the set composed by each combination of a group index in s^+ and a group index in s^-, in our case study, $1170 * 50$, or 585000 possible actions each state.

The feature set $F(s, a)$ of a given state s and action a is given by the equation 4:

$$F_i(s,a) = \begin{cases} 1, \text{if group of index } i \text{ is in } s^+ \text{ given action } a \\ 0, \text{otherwise} \end{cases} \tag{4}$$

Reward

To obtain a reward function for each state s, we used the Menpo Project (Alabort-i-Medina et al. 2014) AAM models and the Lucas-Kanade fitting method (Lucas, Kanade, and others 1981). We train the AAM with the image set composed by the group indexes in s^+, following we test the AAM produced by s^+ against the 60 test cases, producing an error vector e. The error measure we used was the euclidean distance between the fitted landmarks, and those on the ground-truth, normalized by the bounding box of the ground-truth image. This is the default error function in Menpo(Alabort-i-Medina et al. 2014).

To generate the reward score, we then compute the root-mean square(RMS) error of e, to penalize large error discrepancies and generate an error score $\in [0, 1]$. The reward for a given state s is given by the equation 5:

$$R(s) = 1 - RMS(e) \tag{5}$$

Training the Model

We organized the training of the model in episodes. With a random initialization of the weight vector w. For each step in a given episode, we chose a new randomly selected state. The algorithm we used can be seen at 1. During each step the algorithm will choose an action, given a ϵ-greedy policy, and apply the chosen action, producing a state s' by moving images from s^+ and s according to the chose action. The algorithm runs for a pre-determined number of episodes.

Algorithm 1: Q-Learning Dataset selection for AAMs

Random initialization of parameters w
foreach *episode* **do**
  $s \leftarrow$ random state initialization
  foreach *step in episode* **do**
    $a \leftarrow \begin{cases} argmax_{a \in A} Q(s,a), & \text{with probability} 1-\epsilon \\ random_{a \in A}, & \text{with probability } \epsilon \end{cases}$
    $s', r \leftarrow apply\_action(a)$
    $u \leftarrow r + \gamma \max_{a'} Q(s', a') - Q(s, a)$
    foreach $w_i \in w$ **do**
      $w_i \leftarrow w_i + \alpha u \frac{\delta Q(s,a)}{\delta w_i}$
    end
    $s \leftarrow s'$
  end
end

Experimental Results

In this section we report the experiments performed and results obtained when selecting a subset of the input dataset. To test the hypothesis that a linear approximation for a value function is sufficient to select a better performing set of images, we ran two experiments. The input dataset was composed of 1170 groups of image and in each experiment the algorithm was set to find the subset of 50 groups of images which minimized the error in the test set. The first experiment was composed of 40 episodes of 100 steps each, we ran it with an α of 0.05 and the decay of ϵ was of 0.1 of its current value per episode.The second experiment was composed of 130 episodes of 60 steps each and an α of 0.1 and the decay of ϵ was of 0.01 of its current value per episode. After training, each experiment produces a list of 50 image group indexes. The main question we seek to answer with these experiments is whether a linear approximation for a value function is a good fit for an AAM model when it comes to choosing images from a dataset.

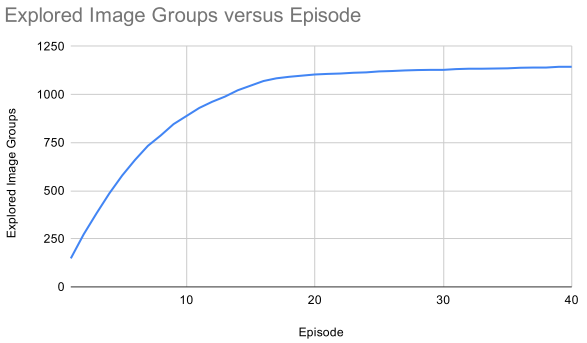
The obtained results can be seen in figure 1 and figure 2. Figure 1 shows the number of explored group indexes, that is, the image groups which were part of any iteration of the algorithm. Figure 2 shows the reward for the state s obtained at the end of each episode.

The first experiment explored a total of 1144 image groups, while the second experiment explored the entire set of 1170 image groups. Neither experiment demonstrated the behaviour of consistently increasing rewards given the entire episode runs.

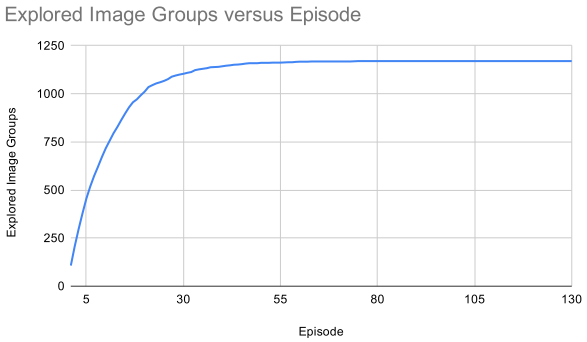
Conclusion

In this section we discuss our obtained results and propositions for future works. In this work we propose a dataset selection algorithm, specifically crafted for Active Appearance Model construction. For this, we shown a Linear Approximation Q-Learning approach. We investigated whether a linear approximation model of Q-learning was the correct abstraction for the relations between images in a dataset used to train an AAM. Our proposed technique was not successful in improving the quality of the image fitting provided by the trained AAM after the entire learning process, using both large and smaller episodes. We can suppose this fact comes from the non-linear relationship of the variance in a facial dataset, how different an image is in a group of images depends of the entire set of images, and not on an intrinsic value of uniqueness.

For future works, we propose the use of a non-linear approximation function for the expected reward of our Q-Learning model, and the use of additional reinforcement learning techniques, such as a more refined state feature function. For a non-linear approximation function, the linear function could be swapped for a neural network without much change to the overall described process, supposedly able to better represent the non-linear nature of difference in a given image set.

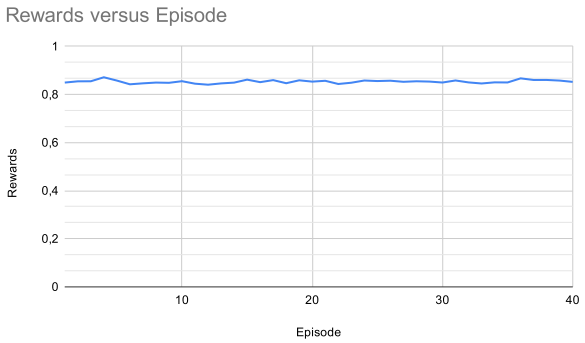


(a) Explored image groups for the first experiment.

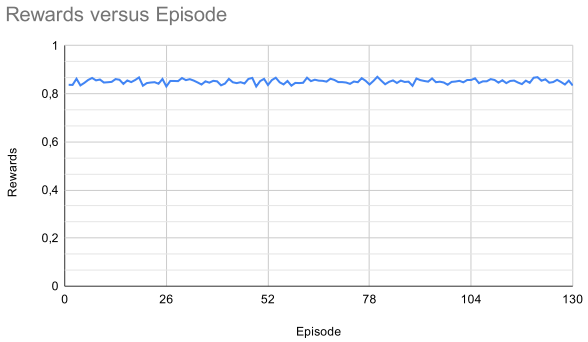


(b) Explored image groups for the second experiment.

Figure 1: The explored image groups at the end of each episode.



(a) Rewards obtained in the first experiment.



(b) Rewards obtained in the second experiment.

Figure 2: The rewards obtained at the end of each episode.

Acknowledgements

We thank Gabriel Licks for insights into which models to investigate.

References

Alabort-i-Medina, J.; Antonakos, E.; Booth, J.; Snape, P.; and Zafeiriou, S. 2014. Menpo: A comprehensive platform for parametric image alignment and visual deformable models. In *Proceedings of the ACM International Conference on Multimedia*, MM '14, 679–682. New York, NY, USA: ACM.

Basu, D.; Lin, Q.; Chen, W.; Vo, H. T.; Yuan, Z.; Senellart, P.; and Bressan, S. 2016. Regularized cost-model oblivious database tuning with reinforcement learning. In *Transactions on Large-Scale Data-and Knowledge-Centered Systems XXVIII*. Springer. 96–132.

Cootes, T. F.; Taylor, C. J.; Cooper, D. H.; and Graham, J. 1995. Active shape models-their training and application. *Computer vision and image understanding* 61(1):38–59.

Cootes, T. F.; Edwards, G. J.; and Taylor, C. J. 1998. Active appearance models. In *European conference on computer vision*, 484–498. Springer.

Cootes, T. F.; Edwards, G. J.; and Taylor, C. J. 2001. Active appearance models. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (6):681–685.

Kaelbling, L. P.; Littman, M. L.; and Moore, A. W. 1996. Reinforcement learning: A survey. *Journal of artificial intelligence research* 4:237–285.

Lucas, B. D.; Kanade, T.; et al. 1981. An iterative image registration technique with an application to stereo vision.

Sutton, R. S., and Barto, A. G. 2018. *Reinforcement learning: An introduction*.

Tsitsiklis, J. N., and Van Roy, B. 1997. Analysis of temporal-diffference learning with function approximation. In *Advances in neural information processing systems*, 1075–1081.

Zhang, Zhifei, S. Y., and Qi, H. 2017. Age progression/regression by conditional adversarial autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.