CS7011 : Topics in Reinforcement Learning
Summary of Overcoming Catastrophic Forgetting : Kirkpatrick, Pascanu, Rabinowitz, Veness

Sahana Ramnath : EE15B109

1. Abstract

This paper (Kirkpatrick et al., 2017) proposes a new method **Elastic Weight Consolidation(EWC)**, analogous to synaptic consolidation for neural networks. In general, neural networks are not capable of learning multiple tasks in sequential order,i.e, continual learning; they are highly prone to *catastrophic forgetting*. It is possible to overcome this limitation and train networks that can remember how to do tasks that they haven't seen in a long time; *EWC* remembers old tasks by selectively slowing down learning on the weights most important to the previously learnt tasks. This approach is scalable and effective. Experiments are conducted on a set of classification tasks based on the MNIST dataset and in reinforcement learning by learning several Atari games sequentially.

2. Continual Learning

In real world settings, an agent will always face multiple different tasks; the sequence of tasks may not be explicitly labelled, the tasks may switch unpredictably and any particular task may not recur for long time intervals. So, agents must show a capacity for *continual learning* : the ability to learn consecutive tasks without forgetting how to perform previously trained tasks. In general, neural networks are not capable of this since when training for a new task B after learning hot to perform task A,all the weights are modified to learn task B. One way to ensure that knowledge of previous tasks remains is to record all data using an episodic memory system and periodically replay them to the network while training; however this is not scalable to a large number of tasks. So, this paper proposes that this problem can be solved by imitating howmammals learn in a continual fashion : task specific synaptic consolidation, where knowledge about how to perform a previously seen task is encoded in a proportion of synapses that are rendered less plastic and so, stable over long timescales. *EWC* attempts to do the same by constraining important parameters for old tasks to stay close to their old values.

3. Elastic Weight Consolidation

In a deep neural network, to learn to perform a task, the weights θ of the network have to be learnt; however, many configurations of θ will result in the same/similar performance. Now, over parametrization makes it likely that there is a solution θ^*_B for task B close to the previously found solution for task A θ^*_A; basically while learning task B, *EWC* preserves the knowledge about task A by constraining θ to stay in a region of low error for task A centered around θ^*_A. This is implemented as a quadratic penalty .

To obtain the importance of a weight with respect to a task, the network has to be viewed from a probabilistic perspective : Finding the optimal θ^* is equivalent to finding the most probable θ given data $\mathcal{D}$. The conditional probability $p(\theta|\mathcal{D})$ is obtained from the prior $p(\theta)$ and the probability of the data given parameters $p(\mathcal{D}|\theta)$ using Bayes' rule : $logp(\theta|\mathcal{D}) = logp(\mathcal{D}|\theta) + logp(\theta) - logp(\mathcal{D})$. Now, $logp(\mathcal{D}|\theta)$ is the negative loss function $-\mathcal{L}(\theta)$. Say the data is split into the two tasks A($\mathcal{D}_A$) and B($\mathcal{D}_B$), with A already learnt. Then : $logp(\theta|\mathcal{D}) = logp(\mathcal{D}_B|\theta) + logp(\theta|\mathcal{D}_A) - logp(\mathcal{D}_B)$. The LHS describes posterior of θ for the entire dataset whereas RHS depends only on the loss for task B. The posterior $p(\theta|\mathcal{D}_A)$ (the prior WRT task B after learning A) must contain information about which parameters were important to A and is approximated as a Gaussian distribution with mean θ^*_A and diagonal precision given by the diagonal of the Fisher information matrix F. F is used for 3 reasons : it is equivalent to the second derivative of the loss near a minimum, it can be computed from first-order derivatives alone and it is guaranteed to be PSD. So now, the objective minimized by *EWC* is : $\mathcal{L}(\theta) = \mathcal{L}_B(\theta) + \sum_i \frac{\lambda}{2} F_i(\theta_i - \theta^*_{A,i})^2$, where $\mathcal{L}_B(\theta)$ is the loss for task B alone, λ represents importance of old tasks WRT new one and i labels the parameter. When starting a third task C, *EWC* will try to keep the network parameters close to the learned θs of both tasks A and B. This can be enforced either with two separate penalties, or as one by since the sum of two quadratic penalties is itself a quadratic penalty.

4. Experiments

4.1. Continual Learning in a Supervised Context

This was done first, to see if EWC could allow deep neural nets to learn a set of complex tasks without catastrophic forgetting : a fully connected multilayer neural net was trained on many supervised learning tasks in sequence; within each task training was done in the traditional way. For each task, a fixed random permutation was generated by which the input pixels of all MNIST images were shuffled making each of them equally difficult as the MNIST problem but each requiring a different solution. Training on this sequence using SGD led to catastrophic forgetting. When the training switches from task A to B, the performance for A declines steeply whereas for B it climbs steeply; the forgetting of A compounds further with time and with adding more tasks. This cannot be solved by regularizing the network with a fixed quadratic constraint for each weight : though the performance in task A declines less rapidly, task B can't be learnt properly as the constraint protects all weights equally, leaving little spare capacity for learning on B; however since *EWC* takes into account the importance of the weight WRT A, it leaves enough capacity for B to learn without forgetting A. To assess whether EWC fits more functionality in the fixed network by allocating separate parts of the network to each task, the overlap of the Fisher matrices of pairs of tasks was calculated; bigger the overlap, more similarity of the 2 tasks and more the weight sharing.

4.2. Continual Learning in RL Context

EWC was then tested in DQNs, in the more demanding RL domain. Previos approaches to continual learning in RL : adding capacity to the network or learning each task in separate networks that are then used to train a single network that can play all of the games. In contrast, EWC makes use of a single network with fixed network capacity and has minimal computational overhead. The agent here used Double Q-Learning with small changes : more parameters, smaller transition table, task-specific bias and gains at each layer, using the full Atari action set, the EWC penalty(Fisher penalty added at each task switch as above) and a task recognition module(task context is obtained from a HMM). The DQN agent was also allowed to maintain separate short-term memory buffers for each task: these allow action values for each task to be learned off-policy using experience replay. The overall system has 2 timescales : over short time-scales, experience replay allows the learning in DQN to be based on the interleaved and uncorrelated experiences and at longer time scales, knowledge across tasks is maintained using EWC. Agent using plain gradient descent methods never learns to play more than one game. By using EWC, however, the agent does learn to play multiple games. The improvement achieved by also providing the task label to the agent was only modest. While augmenting the agent with EWC allows it to learn many games in sequence without catastrophic forgetting, it does not reach the score that would have been obtained by training separate DQNs on each game. This suggests that the algorithm is overconfident about certain parameters being unimportant: so the main limitation of the current implementation is that it underestimates parameter uncertainty.

References

Kirkpatrick, James, Pascanu, Razvan, Rabinowitz, Neil, Veness, Joel, Desjardins, Guillaume, Rusu, Andrei A, Milan, Kieran, Quan, John, Ramalho, Tiago, Grabska-Barwinska, Agnieszka, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, pp. 201611835, 2017.