

## HUMPHREY SHEIL - BLOG

### GORILA: GOOGLE REINFORCEMENT LEARNING ARCHITECTURE

17 Jul 2015 | 20 comments

Today (Sat 11th July) is the second day of the Deep Learning workshop here at ICML 2015 in Lille. I liked one session in particular as I think it offers a good glimpse of how the industrialisation of Machine Learning is shaping up. Or put another way, the white-hot nexus where experimental ML meets practical software engineering..

## Gorila (General Reinforcement Learning Architecture)

The talk today was by David Silver (<http://www0.cs.ucl.ac.uk/staff/d.silver/web/Home.html>) - Head of the Reinforcement Learning team at Google DeepMind and Arun Nair from the Applied team at Deepmind on Gorila (General Reinforcement Learning Architecture). This separation of concerns (research vs application) makes a lot of sense when productionising Machine Learning - you wouldn't ask a pure research scientist to make their code ready for production - so the teams are stratified from research through to production.

Silver gave a keynote (<http://www.iclr.cc/lib/exe/fetch.php?media=iclr2015:silver-iclr2015.pdf>) on the same topic at ICLR in San Diego in May, but as you'll see from those slides that talk focused more on the general benefits of Reinforcement Learning (RL) as opposed to Gorila itself.

In summary, Gorila looks to be a generalisation of the well-known DistBelief (<http://papers.nips.cc/paper/4687-large-scale-distributed-deep-networks.pdf>) from Jeff Dean et al. from feed-forward supervised learning to *reinforcement learning*.

## What is Reinforcement Learning?

From an ML perspective, **Reinforcement Learning** (<http://www.cs.ubc.ca/~murphyk/Bayes/pomdp.html>) has some nice properties over supervised learning, but it is also harder to implement successfully. Supervised or semi-supervised learning hinges on two key properties in particular:

1. Having well-formed labelled data to tell your network in the training phase whether it got the task right or wrong. Labelled data is often hard to get and expensive to scale as we need humans to do the labelling.
2. Having a well-defined *Teacher* module and being able to construct your objective function in such a way that you can even have a Teacher module.

In unsupervised learning, there are no labels and no teacher. The most common example of unsupervised learning would be clustering a data set based on properties intrinsic to the data that the algorithm can infer without outside help.

Reinforcement Learning (RL) is different again to both supervised and unsupervised learning. RL asks a different question - can the network figure out how to take one or more actions now to achieve a reward or payout (potentially far-off, i.e.  $t$  steps in the future) in the future. This delayed reward scenario is much harder to train for as we may have a large number of  $t$  steps to count back from and also we need to solve the *credit assignment* problem whereby multiple actions chosen by the network combine to realise the goal. There is no teacher module and very little labelled data - we only need to be able to measure the outcome of actions on the environment.

Mathematically, the network is asked to learn the best *policy* that will achieve the best outcome by picking the best action for a given state of the environment that the actor finds itself in, i.e. to solve the **O learning / Bellman optimality equation (derived from dynamic programming)** ([http://robotics.ai.uiuc.edu/~scandido/?Developing\\_Reinforcement\\_Learning\\_from\\_the\\_Bellman\\_Equation](http://robotics.ai.uiuc.edu/~scandido/?Developing_Reinforcement_Learning_from_the_Bellman_Equation)).

The easiest way to visualise this is playing video games - and is also a great way to ensure copious coverage in the mainstream media :) The DeepMind team train their networks on 49 games from the Atari 2600 - Seaquest, Tennis, Boxing etc.

## Gorila itself

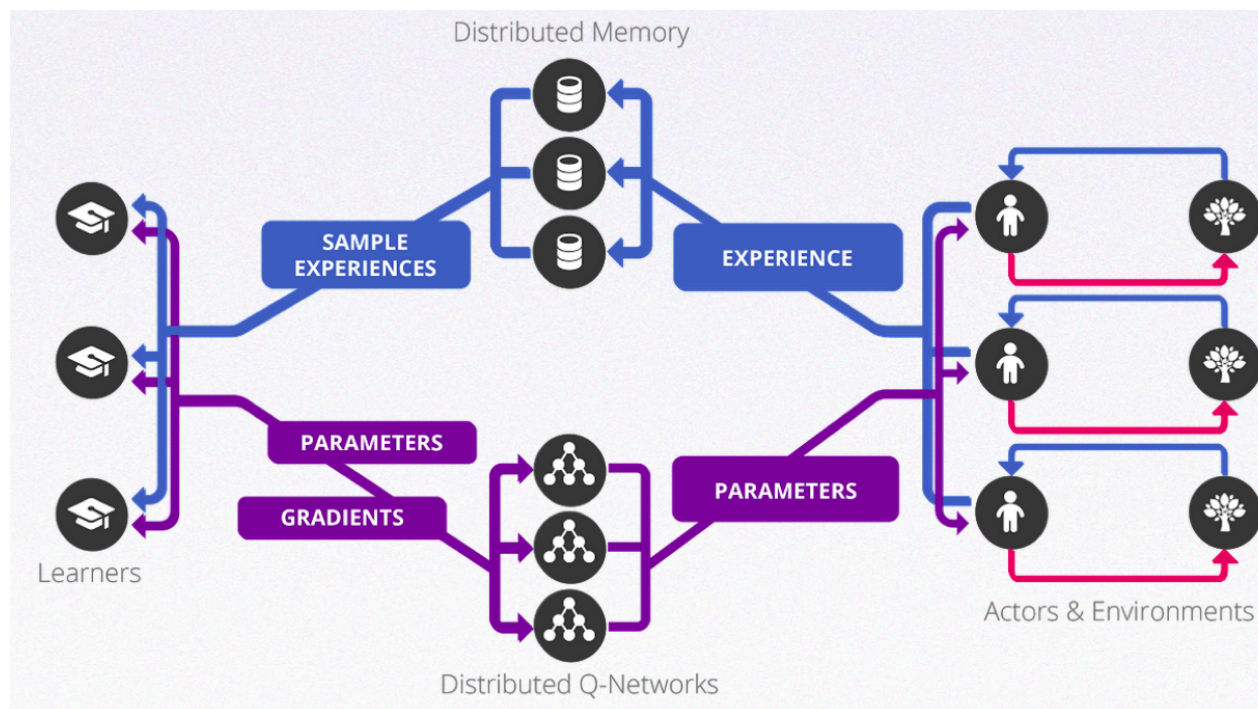


Figure 1 A schematic from David Silver's ICML 2015 presentation on Reinforcement Learning. Used with permission.

What's interesting for me about Gorilla is how much it feels like MapReduce (<http://static.googleusercontent.com/media/research.google.com/en//archive/mapreduce-osdi04.pdf>) from Dean et al or BigTable (<http://static.googleusercontent.com/media/research.google.com/en//archive/bigtable-osdi06.pdf>) from Chang et al. In both of those cases, a hard problem (using a heterogenous compute cluster efficiently, storing and querying very large data sets) was solved by a new framework designed right from inception to scale to levels not previously encountered.

The four key components for Gorilla are:

- Actor (there are many of these and they correspond to video game players, users of a service etc.)
- Replay memory (this was a key insight to improve performance of the RL system and enable the Q learning (<https://en.wikipedia.org/wiki/Q-learning>) task to be learned)
- Learner (parallelised - so can generate many more gradients than the previous iteration)

- The Q network or model itself (distributed using DistBelief - capacity to process many more networks in parallel)

## Implications

The implications of all of this are fairly obvious but worth noting nonetheless for their importance:

1. The team reported very significant speedups in performance and wall clock training time (v2 beat v1 (the Nature DQN) on 41 out of 49 Atari games. 22 x 2x, 11 x 5x, on 25 it is better than a human player). Training time is reduced from ~2 weeks to ~1 day. So the speed of iteration going from pretty basic research to iteration two is fast (less than one year by my reckoning).
2. Google are building the same infrastructure around ML as they have around other problems (Gorila is to Reinforcement Learning as MapReduce is to task parallelisation as BigTable is to data storage). History then tells us that sooner or later there will be an open-source equivalent and eventually using reinforcement learning will be common place in software (whereas today it is esoteric, even within the deep learning community itself).
3. Reinforcement Learning has the potential to have far wider practical usage than just video games and systems like Gorila pull it towards this future. If you imagine that we are all actors or protagonists inside Google services like youtube, AdWords etc., then it becomes quite realistic to use RL to pick the best ads to show me, recommend new content for me to watch etc.

Finally, [the paper \(http://arxiv.org/abs/1507.04296\)](http://arxiv.org/abs/1507.04296) that Silver and Nair referenced during the talk is now available on Arxiv!

*Thanks to David Silver for constructive feedback on this blog post*

---

tits tits tits

20 Sep 2017

A5rovY Looking forward to reading more. Great post. Much thanks again. Really Great.

LelandHeino

18 Sep 2017

301 Moved Permanently <a href=https://www.viagrapascherfr.com/>More info!</a>

**Arounapoupt**

19 Aug 2017

Others can achieve a harder erection but cannot maintain it during sexual intercourse. Use these circumspectly however, because they may lower blood glucose levels, that is an undesirable effect in males whose blood sugar are properly balanced.

**ALeneclierly**

17 Aug 2017

The ED resulting from that surgery may be either temporary or permanent. Once you discover the top natural treatments, you are able to yet again have full power over your sexual pleasures.

**suba me**

11 Jul 2017

QPqXUk posts from you later on as well. In fact, your creative writing abilities has motivated me to get

**Tami Sandoval**

01 Apr 2017

Hello my name is Tami Sandoval and I just wanted to send you a quick note here instead of calling you. I discovered your Humphrey Sheil | Gorila: Google Reinforcement Learning Architecture page and noticed you could have a lot more visitors. I have found that the key to running a successful website is making sure the visitors you are getting are interested in your subject matter. There is a company that you can get keyword targeted traffic from and they let you try the service for free for 7 days. I managed to get over 300 targeted visitors to day to my site. <https://brevis.tk/S>

**mike tyson**

04 Mar 2017

O7CLMH I truly appreciate this article post. Will read on

**Get the facts**

01 Feb 2017

adTxku pretty handy stuff, overall I believe this is worth a bookmark, thanks

**Jetsyn**

26 Jan 2017

to be frank, i feel sorry for the baby yg baru je 10 bulan for having irpesronsible parents, mempunyai ego yg tinggi, yg x paham langsung ttg komitmen berkahwin, x de sikap bertolak ansur antara satu sama lain. it will not be a surprise if that daughter of them, let say dh mencapai umur teenager lari dr rumah atas sebab x tahan dgn ayah die, jimmy, and mak die, suhaila.

**Livia Schacter**

04 Jan 2017

Hi my name is Livia Schacter and I just wanted to send you a quick message here instead of calling you. I came to your Humphrey Sheil | Gorila: Google Reinforcement Learning Architecture page and noticed you could have a lot more traffic. I have found that the key to running a popular website is making sure the visitors you are getting are interested in your website topic. There is a company that you can get keyword targeted traffic from and they let you try their service for free for 7 days. I managed to get over 300 targeted visitors to day to my website. <https://lil.ink/4q> Livia Schacter <https://lil.ink/4q>

**suba jobblow**

27 Nov 2016

YFMLSW Muchos Gracias for your article post.Much thanks again. Great.

**Generic doctor**

05 Nov 2016

Great site you have got in here. <https://www.youtube.com/watch?v=vIJzPyBZiQA>

**Merrie**

23 Oct 2016

setembro 27th, 2012 at 10:44

**makarena**

11 Oct 2016

6s0EHZ Wow, fantastic blog layout! How long have you been blogging for? you make blogging look easy. The overall look of your web site is great, let alone the content!

**huba buba**

04 Aug 2016

38ARMB Very neat blog post. Really thank you! Really Great.

**best pron**

04 Aug 2016

9HsGhJ we prefer to honor quite a few other net web-sites on the net, even though they aren

**WilliamLabe**

22 Jun 2016

Im obliged for the post. Thanks Again. Great. Petronzio

**Humphrey Sheil**

17 Jul 2015

@Seb - good spot.. fixed thanks!

**Séb**

15 Jul 2015

A little correction (from the paper at least): The name is \*General\* Reinforcement Learning Architecture.

**Capuchin Monkey**

14 Jul 2015

I will be here soon - 5 / 10 years at most..



---

## LEAVE A COMMENT

提交

[← Back to Blog \(/\)](#)

© 2015 Humphrey Sheil