Published in Towards Data Science

Debmalya Biswas Follow
Feb 11 · 9 min read ★ · Listen

Data-Driven (Reinforcement Learning-Based) Control

Why Reinforcement Learning is such a hot topic for control-systems optimization problems in general

Fig: Reinforcement Learning (based on the pic by Andrea Piacquadio from Pexels)

Introduction

Data Driven Control, esp. those based on Reinforcement Learning (RL) control strategies, is the new buzzword for Industrial Engineering. RL seems to be the go-to paradigm for all control problems, from controlling combustion engines, to robotic arms cutting metals, to air conditioning systems in buildings. (This is also evident in the number of RL related job advertisements put out by such companies.) We see a similar disruption to how Deep Learning models disrupted traditional Computer Vision techniques for Image Recognition & Segmentation.

> We define Data Driven Control as simply Machine Learning (ML) techniques applied

Control Theory Limitations

At a very basic level (and high-level), a Control System literally consists of a System & Controller:

- *System* to control
- *Controller* applies a control strategy to control the system in an optimal fashion.

There are two other things that we need to consider in this context: Any strategy that the Controller can apply is constrained by

- its knowledge of the system state — in most cases, provided by the System Sensors;
- and the system parameters that it can control — also referred to as the System Actuators. E.g., an engine can only drive a car within a certain speed range, at a certain acceleration.

External/environmental factors also play a key role in the control strategy, however that role is more as an 'input' parameter, and not a constraint. For example, the outdoor temperature play a key role in deciding how much to cool for an air-conditioner; however the air-conditioner's functioning is not constrained by it.

The figure below provides an illustration of a Control System, where x_t is a time derivative of the non-linear function *f*.

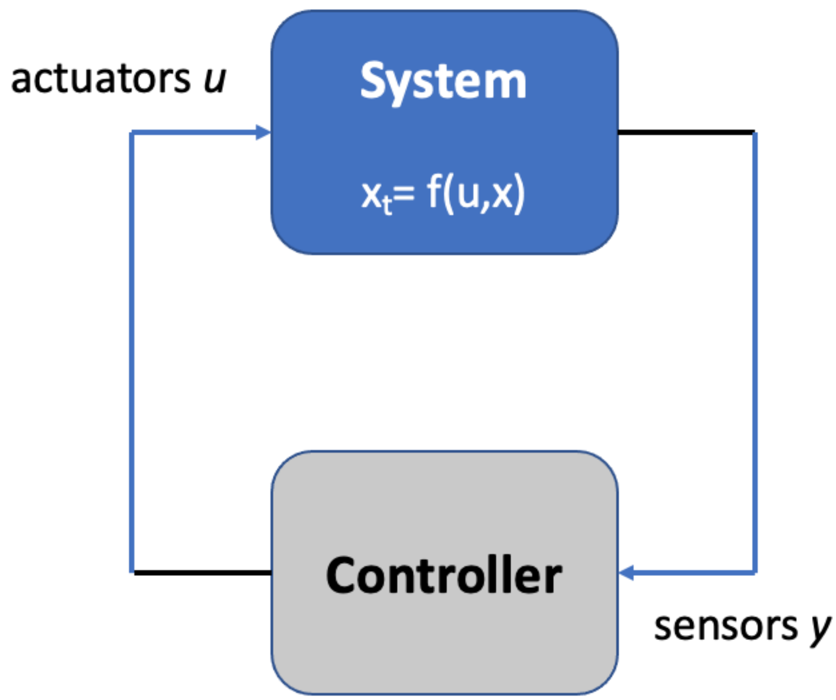

Fig: (Simplified) Control system design (Image by Author)

> Designing a control strategy then consists of solving the equations characterizing the system behavior — often modeled in the form of linear equations. Most of Control Theory is targeted towards solving linear equations.

Unfortunately, real-world systems are (mostly) non-linear. E.g., even the equation to capture the motion of a pendulum is non-linear. There has been a lot of research on linearization methods, basically techniques to convert non-linear equations to linear ones and then systems that we do not know how to model (whose system equations are not known). And, the complexity of such systems is only increasing day by day, where we want to solve hyper-scale problems, e.g., climate control, disease control, automated vehicles, financial markets, etc.

To summarize the limitations of traditional Control Theory/Model Driven Control:

- System models/equations are not known
- Do not work for large-scale non-linear domains
- Simulation of such systems are also very difficult given their high dimensionality

For a detailed discussion on this topic, refer to Steve Burton's excellent tutorial on Control Systems [1].

ML to the Rescue

Given the above challenges with traditional Control Theory, let us now try to understand why ML/Data based approaches show a lot of promise in this context.

> The underlying logic here is that even for a very high dimensional system that we cannot model, there are dominant patterns that characterize the system behavior — and Machine Learning (Deep Learning) is very good at learning these patterns.

This would (most likely) be an approximation, and while we still would not understand the system fully — it is good enough for most real-life use-cases (including predictions), barring some exceptional scenarios.

In this article, we focus on RL based approaches for Control Systems. We will also touch upon the key differences between supervised ML and un/semi-supervised RL, and how this makes RL a good choice for potentially any control optimization problem.

Reinforcement Learning (RL)

RL is able to achieve complex goals by maximizing a reward function in real-time. The reward function works similar to incentivizing a child with candy and spankings, such that the algorithm is penalized when it takes a wrong decision and rewarded when it takes a right one — this is reinforcement. The reinforcement aspect also allows it to adapt faster to real-time changes in the user sentiment. For a detailed introduction to RL frameworks, the interested reader is referred to [2].

- *Rewards and Policies are not the same*: The roles and responsibilities of the Reward function vs. RL Agent policies are not very well defined, and can vary between architectures. A naïve understanding would be that given an associated reward/cost with every state-action pair, the policy would always try to minimize the overall cost. Apparently, it seems that sometimes keeping the ecosystem in a stable state can be more important than minimizing the cost (e.g. in a climate control use-case). As such, the RL Agent policy goal need not always be aligned with the Reward function, and that is why two separate functions are needed.

- Similar to supervised approaches in Machine Learning/ Deep Learning, the *RL approach most suitable for enterprise adoption is 'Model based RL'*. In Model based RL, it is possible to develop a model of the problem scenario, and bootstrap initial RL training based on the model simulation values. For instance, for energy optimization use-cases, a blueprint of the building Heating, Ventilation and Air Conditioning (HVAC) system serves as a model, whose simulation values can be used to train the RL model. For complex scenarios (e.g. games, robotic tasks), where it is not possible to build a model of the problem scenario, it might still be possible to bootstrap an RL model based on historical values.

> This is referred to as 'offline training', and is considered a good starting point in the absence of a model. And, this is also the reason why RL is often considered as a hybrid between supervised and unsupervised learning, rather than a purely unsupervised learning paradigm.

- *Online and model-free RL remain the most challenging*, where the RL agent is trying to learn and react in real-time without any supervision. Research in this field seems to lack a theoretical foundation at this stage. Researchers are trying out different approaches by simply throwing more data and computing power at the problems. As such, this remains the most "interesting" part of RL, with current research primarily focusing on efficient heuristics and distributed computation to cover the search space in an accelerated fashion. Applying DL (neural networks) to the different RL aspects, e.g., policies, rewards, also remains a hot topic — referred to as Deep Reinforcement Learning [4].

- Given the fundamental nature of RL, there seems to be many interesting concepts that can be borrowed from existing research in *Decision Sciences and Human Psychology*. For example, an interesting quote from Tom Griffiths, in his presentation "Rational use of cognitive resources in humans and machines" [5]:

> while mimicking the human brain seems to be the holy grail of AI/RL research; humans have long have been considered as essentially flawed characters in psychological studies. So what we really want to do is of course to mimic the "rational behavior" of the human brain.

The summary is of course that we need to bring the two fields together if we ever want machines to reach the level of true human intelligence.

Case Study: RL based HVAC Optimization

D. Biswas. *Reinforcement Learning based Energy Optimization in Factories*. (Towards Data Science — link), also published in proceedings of the 11th ACM e-Energy Conference, Jun 2020.

> The above article is an interesting case study in the context of our current discussion. It showcases the successful transition of an Industrial Control System run by a traditional PID controller for the last 10+ years, to a more efficient RL based controller.

The Industrial Control System in this case refers to Heating, Ventilation and Air Conditioning (HVAC) units responsible for maintaining

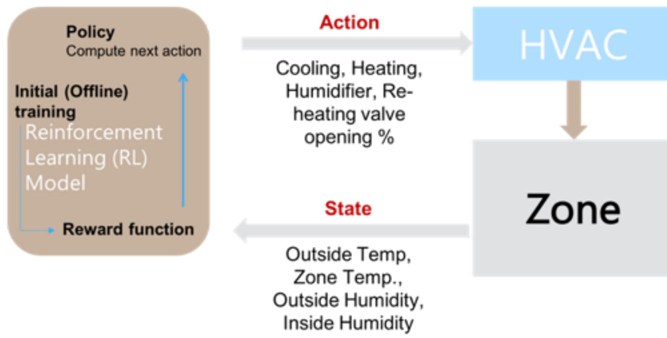

Fig: HVAC Reinforcement Learning formulation (Image by Author)

This is a complex problem as it requires computing an optimal state considering multiple variable factors, e.g. the occupancy, manufacturing schedule, temperature requirements of operating machines, air flow dynamics within the building, external weather conditions, energy savings, etc.

Initial state: when we initially started exploring the feasibility of RL based approaches to optimize the HVAC controllers, the HVAC units were controlled by PID controllers. PID (proportional integral derivative) [8] is a popular control technique for optimization problems, which uses a control loop feedback mechanism to control process variables.

This had been the case for the last 10+ years, and the HVAC units were doing their job of maintaining the factory temperature and humidity at their desired settings in a very stable & robust fashion. So the RL based exploration was primarily motivated by their potential to reduce energy consumption and CO_2 emissions. (Studies have shown that HVAC accounts for almost 50% energy consumption in a building and 10% of global electricity usage.)

Solution: Given the difficulty of modeling such systems, we started with an initial deployment of a stochastic RL model (in the sense that it only depends on the last state). In the next stage, we extended the RL model to accommodate 'long term rewards', quantified by the Q-value in RL terminology. Q-value for a state-action pair *(s, a)* is defined as a weighted sum of the expected reward values of all future steps starting from the current state *s*, given that action *a* is taken at state *s*. This RL model operates in a continuous space setting. Each episode in this setting corresponds to a period when the indoor temperature and (or) humidity starts moving away from their respective setpoints, to the time that the indoor conditions return to their respective setpoint values — as a result of opening the relevant valve(s).

Target state (ongoing): Within a 6-month pilot, we were able to develop and operationalize a RL based HVAC controller that is able to learn and adapt to real-life factory settings, without the need for any offline training. Benchmarking results show the potential to save up to *25%* in energy efficiency (as compared to when they were operated by PID controllers).

Conclusion

To conclude, RL has had an interesting journey so far. From the hype generated by unsupervised RL agents beating AlphaGo players, to struggling to find a place/utility in the enterprise world. It has been a similar undulating journey on the research side as well, with interest in RL models peaking in the last few years as progress in Deep Learning models saturated; to now the focus again shifting to Self-supervised Systems.

Having said this, RL seems to have found a sweet spot in Industrial Control Systems. There has been some progress in applying RL techniques to Recommenders [3], Chatbots [9]; however control optimization is where they 'best fit'. In this article, we highlighted the challenges of traditional Control Theory, and made the point for RL based Controllers to potentially solve/improve the many complex problems of this domain. It is an exciting time to get involved in this journey and hope you feel the same —looking forward to your feedback !

References

[1] Steve Brunton: *Control Bootcamp*, https://www.youtube.com/watch?v=Pi7l8mMjYVE&list=PLMrJAkhIeNNR20Mz

[3] D. Biswas. *Reinforcement Learning based Recommender Systems.* (Medium link — Towards Data Science), also presented in the ""Advances in Artificial Intelligence for Healthcare" track at the 24th European Conference on Artificial Intelligence (ECAI), Sep 2020.

[4] W. Dabney: *Advances in Distributional Reinforcement Learning and Connections With Planning,* 2020, https://www.youtube.com/watch?v=iqIGHSgYtbs

[5] T. L. Griffiths, F. Lieder, N. D. Goodman. *Rational Use of Cognitive Resources: Levels of Analysis Between the Computational and the Algorithmic.* https://cocolab.stanford.edu/papers/GriffithsEtAl2015-TiCS.pdf

[6] F. Oldewurtel and et al. *Energy efficient building climate control using stochastic model predictive control and weather predictions*. ACC, 2010.

[7] Y. Ma and et al. *Model predictive control for the operation of building cooling systems*. IEEE Transactions on Control Systems Technology, 20(3):796–803, 2012.

[8] F. Peacock. *An Idiot's Guide to the PID Algorithm*. https://www.pidcontrol.net/index.html

[9] Ricciardelli, E., Biswas, D.: *Self-improving Chatbots based on Reinforcement Learning*. (Medium link — Towards Data Science) In: 4th Multidisciplinary Conference on Reinforcement Learning and Decision Making (2019).