德州農工大學 P. R. Kumar 教授演講：

From Adaptive Control to Reinforcement Learning

文／洪鈺恒　資科工博士生

強化學習可視為機器學習的一個子領域，不同於監督式學習 (Supervised learning) 和非監督式學習 (Unsupervised learning)，強化學習重點在控制探索未知領域 (Exploration) 和利用現有對環境的認識 (Exploitation) 的平衡。可以認為機器學習的目的是要學習資料 (data) 和標籤 (label) 之間的關係，亦可稱之為回歸 (Regression)，而強化學習則是要一步步的跟環境互動來產生學習需要的資料，而學習的過程中也要最大化從環境中得到的獎勵，為了達到此目的就需要一個好的策略來控制要 " 探索未知 " 還是 " 選取當下最好的動作 " 的比例。

強化學習可視為機器學習的一個子領域，不同於監督式學習 (Supervised learning) 和非監督式學習 (Unsupervised learning)，強化學習重點在控制探索未知領域 (Exploration) 和利用現有對環境的認識 (Exploitation) 的平衡。可以認為機器學習的目的是要學習資料 (data) 和標籤 (label) 之間的關係，亦可稱之為回歸 (Regression)，而強化學習則是要一步步的跟環境互動來產生學習需要的資料，而學習的過程中也要最大化從環境中得到的獎勵，為了達到此目的就需要一個好的策略來控制要 " 探索未知 " 還是 " 選取當下最好的動作 " 的比例。比較著名的應用為遊戲的人工智慧，像是 AlphaGo 的圍棋 AI，機器人控制以及推薦系統的演算法。而控制領域被認為是強化學習的前身，主要是在研究一個動態系統的行為，傳統是透過工程數學的方法來設計可以讓動態系統穩定的控制器。用這謝教授的話來概括強化學習就是學習式的控制系統 (Learning-based control)。

這次的演講主要分為兩個段落，第一個段落 Kumar 介紹了控制系統中的核心 - 自調諧調節器 (self-tuning regulators) 以及他的四個主要理論性質，分別為 stability，self-optimality，self-tuning 和 strong consistency，相關的理論證明持續了將近 30 年，直到西元 1980 年代才被解決，Kumar 只用短短的兩小時就能讓我們了解這 30 年這些理論的發展過程以及細節。在這個段落的最後 Kumar 則介紹了控制領域中更廣泛討論的情況 -adaptive controllers for armax systems。

介紹完控制領域的背景以及理論後，第二段的主題則是這次演講的題目 - 從自適應控制到強化學習，Kumar 從最簡單的強化學習的範例開始 - 多臂老虎機問題 (multi-armed bandit problem)，介紹最簡單的強化學習演算法的同時也連結了強化學習和控制領域的架構，這些簡單的例子使我們更容易以理論的角度來討論這兩個領域。最後，Kumar 也提到了在控制領域中一個很經典的問題是 "Closed-loop identification"，大意是說在 Maximum likelihood estimation (MLE) 的學習方式下會因為缺乏探索而無法學到真正的系統參數，而 Reward-biased maximum likelihood estimation (RBMLE)” 則是用來解這問題的一個很經典的演算法，可以想成是在 MLE 上加上一些擾動來達到探索的目的，講者也分享了 RBMLE 近年來被用來解強化學習問題的幾篇論文。這些研究使我們瞭更加了解如何把強化學習聯想成是控制領域的問題來解決。

演講結束後， 這次演講的主辦人 - 謝秉均教授也與 Kumar 討論了幾個和 RBMLE 有關的問題，像是如果 likelihood 本身沒有凹 (convex) 的性質以及用梯度下降的方式來解 RBMLE 的情況，這些問題都使我們更了解 RBMLE 這個演算法的概念。此外，謝教授也請 Kumar 提供了幾個建議給剛踏進強化學習領域的人如何開始研究與學習。這次的演講以及這些建議都讓我們收穫許多。

P. R. Kumar’s Speech, “from Adaptive Control to Reinforcement Learning”

P. R. Kumar is currently a distinguished professor at Texas A & M University. His research interests include game theory, adaptive control, machine learning, power systems, automated transportation, unmanned aerial traffic management, millimeter wave, and cyber-physical systems. In this UI NYCU AI lab event, he gave a talk on investigating the relationship between reinforcement learning and control theory.

Reinforcement learning is a sub-domain of machine learning. Different from supervised learning and unsupervised learning, it focuses on controlling the balance between exploration and exploitation of existing knowledge of the environment. It can be said that the purpose of machine learning is to learn the relationship between data and labels, which is also called regression. Reinforcement learning is to interact with the environment step by step to generate the data needed for learning. The learning process also needs to maximize the rewards from the environment. Therefore, to achieve this purpose, it requires a good strategy to control the trade-off of exploring the unknown or choosing the best action of the moment.

The famous AI applications of reinforcement learning are AlphaGo, robot control, and recommender system algorithms. Control theory is seen as the predecessor of reinforcement learning, which is the study of the performance of a dynamic system. Traditionally, stable controllers are designed through engineering mathematical methods. “Reinforcement learning is a kind of learning-based control”, quoted from Prof. Hsieh. This lecture mainly focused on two parts. In the first session, Dr. P. R. Kumar introduced the core part of the control system, which are self-tuning regulators, and the four main theories of the dynamic system, including stability, self-optimality, self-tuning, and strong consistency. These theoretical proofs lasted for almost 30 years and were not solved until the 1980s. It was very impressive that Dr. P. R. Kumar only spent 2 hours giving us a comprehensive insight into the development of these theories over the past 30 years. At the end of the session, Dr. P. R. Kumar introduced some more widely discussed cases, which were adaptive controllers for Argmax systems.

After introducing the background and theory of adaptive control theory, the second part of the lecture focused on the speech title, From Adaptive Control to Reinforcement Learning. He started from a classical problem, the multi-armed bandit problem to connect with reinforcement learning and adaptive control models. Through these examples, we could understand these two areas from a theoretical perspective. Dr. P. R. Kumar also mentioned another classical question, closed-loop identification in adaptive control theory, which referred to the learning method of Maximum Likelihood Estimation (MLE), which may fail to learn the real system parameters due to lack of exploration. On the other hand, Reward-Biased Maximum Likelihood Estimation (RBMLE) is a classic algorithm for solving this problem. It can be thought of as adding some disturbance to the MLE to achieve the purpose of exploration. The speaker also shared several papers in which RBMLE has been used to solve enhanced learning problems in recent years. These studies gave us a better understanding of how to associate the problems in reinforcement learning with control theory.

After the talk, the host of this event, Prof. Ping-Chun Hsieh, discussed several issues with Dr. P. R. Kumar related to RBMLE, such as the quality of the likelihood itself if it is not concave and the case of solving RBMLE by gradient descent. In addition, Prof. Hsieh asked Dr. P. R. Kumar to provide several suggestions on how to start research and learning for those who are new to reinforcement learning. Overall, we learned a lot from this talk and these suggestions.