

# 以強化式學習達到機械手臂避障及能量速度 優化之軌跡規劃

## Manipulator Trajectory Planning for Obstacle Avoidance and Energy/Speed Optimization Based on Reinforcement Learning

陳永修<sup>1</sup>、林沛群<sup>2</sup>

<sup>1</sup>國立臺灣大學 機械工程學系 碩士研究生

<sup>2</sup>國立臺灣大學 機械工程學系 教授

主要聯絡人的電話及mail

林沛群 peichunlin@ntu.edu.tw

**摘要：**近年來工業自動化為發展趨勢，而自動化仰賴機器人的應用。隨著機器人的應用日益增加，則機器人運作效率要求也漸增。本研究著重於強化式學習優化軌跡，並以機器學習法補償動力學模型。先建立機械手臂的動力學模型，以機器學習法補償模型與真實系統的差異，接著，軌跡優化則以教導模式生成軌跡以決定的中途點，再透過強化式學習，調整中途點，來達到能量最佳化或者時間最佳化。並且以實驗驗證此方法之可行性。

**Abstract :** As the development of industrial automation, robot manipulators are widely applied. Since the reliance on robot arms increases, demands on performance of robot arms also rises. This research focused on energy- and time-optimal trajectory planning by reinforcement learning method. At first, dynamic model of a robot arm was established. And machine learning method was used to compensate errors between model and real system. Then, via points of an initial trajectory could be obtained by teaching method. Reinforcement learning method could minimize energy consumption and time duration along the trajectory. The results were verified by experiments.

**關鍵詞：**機器學習、軌跡優化、強化式學習

**Keywords :** Machine learning, Trajectory optimization, Reinforcement learning

### 前言

隨著工業自動化的發展，由於機器沒有疲勞問題，所以機器逐漸取代人力。其中機器手臂的應用日漸增加，因為機械手臂與人類手臂類似，可以執行相同任務，且生產線不需要做變更，然而，機械手臂的應用日漸廣泛，其性能的要求也跟著提升。軌跡的能量優化可以減少機械手臂的耗能，若機器人在生產線上數量增加，能量消耗為重要議題，軌跡能量的優化有其必要性。由於工作效率的需求，軌跡時間優化為另一議題，在重複作動的情況下，軌跡運作時間縮短，則工作效率能提升。因此，本

研究著重於軌跡能量與時間優化。

一般處理機械手臂軌跡能量優化的問題，設定Cost Function以能量消耗為目標[1-3]，使機械手臂減少能量消耗。在時間優化方面，傳統上僅考慮簡化的三軸機械手臂[4-6]，來進行軌跡時間優化。或者，考慮簡化的三軸機械手臂[7, 8]，進行力矩優化，使力矩小於特定上限。其他研究需要複雜的數學推導，或者是需要制定合適的Cost Function，所以本研究採用強化式學習[9]，藉由獎勵高低的方式讓機械手臂學習，軌跡能量或時間越少，則分數越高，達到使機械手臂的軌跡能量或時間最佳化，免去複雜的數學演算或是Cost Function的制定。

► 以強化式學習達到機械手臂避障及能量速度優化之軌跡規劃

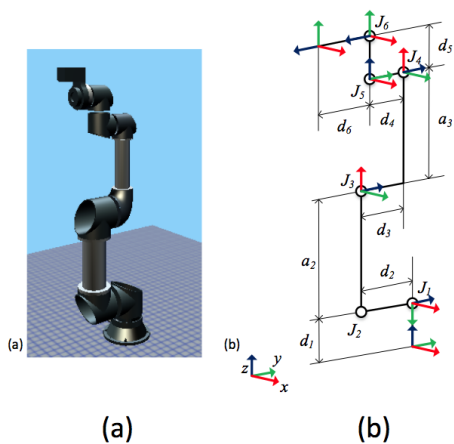


圖1 TM5-900 (a) 手臂機構 (b) 座標系統

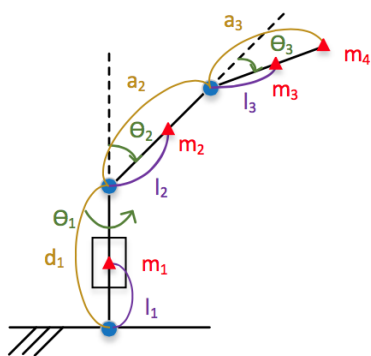


圖2 簡化TM5-900模型

### 動力學模型建立

本研究平臺為達明機器人公司生產TM5-900機械手臂，如圖1，具有六軸自由度。

目前常用的工業機械手臂，多為六軸機構，考量後三軸馬達功率較小，不若前三軸需負載後三軸重量，且機械手臂移動時，後三軸慣性力及科氏力影響整體能量消耗相對小，所以建立動力學模型以前三軸為主，末三軸則視為一個質點圖2。以Lagrange Equations方式建立模型，如式(1)

$$\tau_d = M(\theta)\ddot{\theta} + C(\theta, \dot{\theta})\dot{\theta} + G(\theta) \quad (1)$$

除了動力學模型以外，馬達傳動系統如馬達慣量或者摩擦力等，亦會影響機械手臂運動的能量消耗，完整的動力學模型如式(2)，、和由機械手臂

$$\tau_i = M(\theta)\ddot{\theta} + C(\theta, \dot{\theta})\dot{\theta} + G(\theta) + [B_m\ddot{\theta} + C_m\dot{\theta} + f_c(\text{sign}(\dot{\theta}))] \quad (2)$$

實驗，將實驗蒐集的和，用線性回歸計算所得，但測試結果模型的誤差仍然偏高，因此採用Neural Network補償動力學模型。Neural Network架構如圖3，輸入為前三軸的 $\theta$ 、 $\dot{\theta}$ 、 $\ddot{\theta}$ ，共9個狀態，中間的hidden layer有5層，輸出則為前三軸的補償力矩。

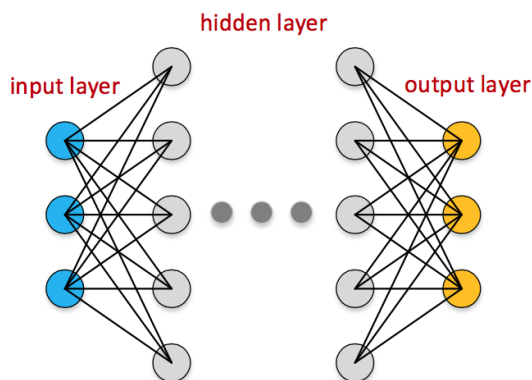


圖3 類神經網絡架構

### 軌跡能量優化

軌跡生成以三次式(cubic spline)生成，給定中途點，用三次式將起點、中途點及終點連接起來，其中起點與終點角度已知，角速度皆為0，中途點的角度已知，角速度部分，若兩段速度同向，取平均速度，若反向，則設為0。

針對是先給定的軌跡，由三次式計算出整段軌跡，用強化式學習來處理軌跡能量優化問題。強化學習的簡易流程如圖4，Agent主要包含了Actor和Critic兩個類神經網路。Actor會產生Action，透過

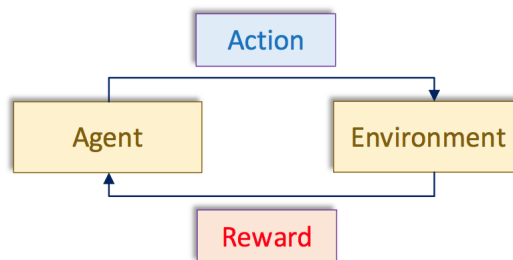


圖4 強化學習簡易流程圖

表1 模擬結果

	中途點	能耗[J]	效果
原始	(60.0°, 10.0°, -110.0°), (70.0°, 20.0°, -120.0°)	51.8016	
暴力搜索	(60.5°, 12.5°, -109.5°), (73.0°, 25.0°, -98.0°)	46.0328	-11.1%
強化學習	(57.5°, 10.5°, -110.8°), (70.9°, 23.5°, -97.3°)	46.5320	-10.2%

表2 教導模式與強化學習優化結果統計資料

	(a)	(b)	(c)	(d)	(e)
原始[J]	36.4549	62.8508	37.2037	52.8886	33.8607
學習後[J]	32.3905	35.5844	32.5055	46.2620	32.1811
改善率	-11.1%	-43.4%	-12.6%	-12.5%	-5.0%

表3 時間優化後之數據

	(a)	(b)	(c)	(d)	(e)
原始[sec]	1.3107	2.0480	1.4418	2.2528	1.3107
學習後[sec]	0.9999	1.6176	1.0822	1.7032	1.0352
改善率	23.7%	21.0%	24.5%	24.4%	21.0%

$$reward = C \times -\frac{1}{dist_{min}} + (1-C) \times \frac{10^4}{\sum Energy} \quad (3)$$

$$\sum_{i=1}^3 \int_{t_i=0}^{t_i=t} \tau_i(t) \dot{\theta}_i(t) dt \quad (4)$$

訓練Actor Network使得它的表現更好，能夠獲得更高的Reward。而Critic則是被用來對Actor打分數，透過學習環境和獎勵之間的關係來調整給分數的策略，也就是更新類神經網路的參數。簡單來說，一開始Actor會隨機動作，Critic隨機給分數，但由於Reward的關係，在訓練的過程中，輸出的Action會逐漸往Reward較高的方向發展。能量優化的Reward如式3，若軌跡會撞到障礙物則C為1，若不會碰撞則為0，表示手臂與障礙物包圍盒的距離，避免能量優化過程產生碰撞，則為手臂消耗能量。

軌跡能量計算方式為整段軌跡的力矩與角速度積分如式4。以暴力搜索法，搜尋全域的能量最佳化軌跡作為比較，實驗以兩個中途點軌跡進行，結果如表1，暴力搜索法及強化學習法皆能節省約10%能量，但是暴力搜索耗時12小時，強化學習僅3分鐘，其運算效率大為提升。

實驗以兩個中途點生成之軌跡表2，模擬時，暴力搜索全域的能量最佳化軌跡與強化式學習結果

相近，能量接減少10%。實際實驗時，以教導機械手臂走五條軌跡，從起始點到終點，中間經過兩個中途點，以及強化式學習結果比較，發現強化式學習能降低能量消耗最佳的情況為節省約40%。

### 軌跡時間優化

軌跡時間優化之目的，在於機械手臂運動速度增快，軌跡運行時間降低，但是由於軌跡速度變快，也代表著軌跡運型的加速度變大，機械手臂力矩增加，則能量消耗也跟著增加，因此設定一力矩上限。強化學習如軌跡能量優化，而reward則如式5。

$$reward = D \times -|\tau_{limit} - \tau_{peak}| + (1 - D) \times (10^3 \times time_{reduction}) \quad (5)$$

若力矩大於上限，則D為1；若力矩小於上限，則D為0。為力矩上限，為力矩峰值，代表軌跡運行時間的減少。當分數越高，則時間優化效果越好。

以軌跡能量優化所得之軌跡為基準，進行時間優化模擬，原始能量優化之軌跡耗時為1秒，經過時間優化以後，耗時降至0.6298秒，時間優化37%

。以教導模式軌跡進行時間優化結果如表3，實驗結果證明，各軌跡耗時皆減少約20%。

## 結語

本研究致力於著重於機械手臂軌跡能量與速度優化，先建立機械手臂的動力學模型，並且以類神經網絡補償模型，使其接近真實機械手臂，來模擬機械手臂原始軌跡能量與時間消耗，與強化式學習能量與時間優化軌跡比較，結果顯示學習後可以降低手臂能量消耗及提升速度。再以實驗結果驗證，教導模式下產生的五條軌跡進行優化，能量消耗可改善約40%與耗時能改善約25%。

## 參考文獻

- [1] A. R. Hiraikawa and A. Kawamura, "Trajectory generation for redundant manipulators under optimization of consumed electrical energy," in Industry Applications Conference, 1996, Vol. 3, pp. 1626-1632.
- [2] A. R. Hiraikawa and A. Kawamura, "Trajectory planning of redundant manipulators for minimum energy consumption without matrix inversion," in IEEE International Conference on Robotics and Automation (ICRA), 1997, Vol. 3, pp. 2415-2420.
- [3] C. Hansen, J. Öltjen, D. Meike, and T. Ortmaier, "Enhanced approach for energy-efficient trajectory generation of industrial robots," in IEEE International Conference on Automation Science and Engineering (CASE), 2012, pp. 1-7.
- [4] G. Sahar and J. M. Hollerbach, "Planning of minimum-time trajectories for robot arms," The International journal of robotics research, Vol. 5, no. 3, pp. 90-100, 1986.
- [5] A. Gasparetto and V. Zanotto, "Optimal trajectory planning for industrial robots," Advances in Engineering Software, Vol. 41, no. 4, pp. 548-556, 2010.
- [6] M. Ghasemi, N. Kashiri, and M. Dardel, "Time-optimal trajectory planning of robot manipulators in point-to-point motion using an indirect method," Pro-

ceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, Vol. 226, no. 2, pp. 473-484, 2012.

[7] J. Hollerbach and K. Suh, "Redundancy resolution of manipulators through torque optimization," in IEEE International Conference on Robotics and Automation (ICRA), 1985, Vol. 2, pp. 1016-1021.

[8] K. Suh and J. Hollerbach, "Local versus global torque optimization of redundant manipulators," in IEEE International Conference on Robotics and Automation (ICRA), 1987, Vol. 4, pp. 619-624.

[9] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," The International Journal of Robotics Research, Vol. 32, no. 11, pp. 1238-1274, 2013.