

學習型系統於人機合作模式下衝突化解機制之研究

計畫編號：NSC 88-2213-E-029-005

執行時間：87年8月1日至87年7月31日

主持人：王偉華 東海大學工業工程系 副教授

摘要

人腦特有的創造力與適應學習能力，是迄今任何電腦系統皆無法達到的。然而，電腦系統快速的計算能力與龐大的儲存能力，亦令人腦望其項背。有效結合人腦與電腦之優點，使人與機器分工合作以解決複雜的大問題，乃人機合作主要的目的。然而，大多數有關人機合作之研究，皆把決策主導地位置於電腦系統或使用者之上，僅做到人與機器的單向支援而非雙向合作。本研究嘗試提出一個使用者與系統居於平等地位之人機合作模式，根據系統與使用者彼此的表現，使決策主導地位能動態地移轉，期望能藉此確實做到人與機器的雙向合作。

本研究以網路表現法表現系統知識，並提出以 Q-Learning 演算法為基礎之重點式增強學習方式，作為學習系統的學習機制。經由法則確定因子與法則門檻值之設計，透過衝突化解機制之運作，使系統與使用者在溝通協調的過程中，決策主導地位能動態地移轉，並使雙方決策能漸趨一致；最後本研究以路徑尋找問題進行實驗，觀察系統行為是否能符合設計時所預期的效果。

關鍵詞：人機合作、增強學習技術、信度分配技術、衝突化解機制

1、前言與相關研究回顧

隨著資訊科技的日新月異，利用電腦系統來輔助甚至取代我們生活中一些工作，以減輕我們的負擔，同時增加工作效益，已經不再是不可能的事。然而，由於真實問題的複雜度以及動態性，除了在一些高度穩定的環境中外，大部分的問題都需要人類的知識擷取以及判斷的能力，才能產生預期的效果。因此使人與機器相互合作，各自發揮所長，以有效解決問題，是將電腦系統引進原本完全由人負責的工作環境之主要研究方向。本研究的目的即在嘗試結合人類與電腦系統的優點，研究人機共同合作來解決問題的可行性。

若人與機器間決策有衝突時，該如何處理？是否能透過有效的處理方法，使得衝突的經驗可增進人機合作系統未來的效能？將來遇到相同情況

時，不再發生衝突？這是在發展人機合作模式系統時，必須周詳考量與解決的問題。在蔡佩芳 [13] 中，便曾探討在人機合作模式下衝突化解機制的設計與運作。其研究中所設計的衝突化解機制，是根據系統的表現績效，以機率性選擇的方式，決定採取使用者或系統的決策。然而、現實環境中需要作決策的狀況，通常需要較長的推理過程。當衝突發生時，人們通常是根據對自己決策的信心，決定是否堅持。同理，人與機器發生衝突時，系統根據對決策的信心，以決定性的方式決定是否堅持，似乎比機率性的方式來的合理。因此，如何以較合理、較符合現實環境的方式來處理人與機器間的衝突，為本研究最主要的動機。

日本鋼鐵 NKK 的煉鋼系統 Scheplan[5] 這個系統主要是一個限制式滿足技術的應用，將煉鋼時資源、設備等的限制條件轉換為系統中的限

制式。但是在發展人機合作排程系統時，發展者將原限制滿足系統，轉換成一個專家系統，將限制式視為法則，建立法則庫，再發展出 Scheplan 的架構。在這個架構下，有三個主要的元素：執程序 (procedures)、法則 (rules) 及使用者 (user) 之間的相互合作。在衝突處理上，Scheplan 系統是利用排程引擎 (Scheduling Engine) 的運作，計算出符合限制條件的合理解 (feasible solutions)，若使用者對於此合理解並不滿意，則利用使用者介面 (User Interface) 來修改此合理解。當使用者所輸入的限制會產生排程上的衝突時，推理引擎會以使用者的輸入為主，自動將不合理的部分修正。在此系統中，使用者居絕對主導地位，最終的決策皆為使用者同意的決策。

還有一個例子是法國的飛機零件製造公司 Dassault Aviation 的排程系統 MADE[3]，為了因應零件製造時，有許多手工加工的製程，所以加工排程必須經過師傅的同意。因此，不同的師傅對加工過程可能有不同的需求，對排程結果便有不同的考量。這個系統主要的目的，是讓使用者在不違背現有的排程限制法則之下，很容易的修改排程結果，以符合個人的要求。當使用者輸入個別的限制時，系統會在符合原生產限制下，再考慮使用者的限制。一旦有所衝突，則以系統的生產限制為主。因此，此系統是以系統居主導地位。

大多數有關人機合作的研究，皆把主導地位置於電腦系統或使用者上。當有決策衝突的情況發生時，總是採用居於主導地位者的決策。因此，只做到單方面的溝通，並未做到讓電腦系統與使用者居於平等地位來溝通協調，並根據彼此對問題的解決能力、掌握資訊與過往經驗來決定採取何者的決策。嚴格來說，這些研究只做到人與機器的單向支援而非雙向合作。因此，本研究嘗試提出一個使用者與

系統居於平等地位之人機合作模式。根據系統與使用者彼此的表現，使決策主導地位能動態地移轉，期望能藉此確實做到人與機器的雙向合作。

2、角色平等人機合作系統設計

根據研究目的，本研究所處理的問題，可歸納出下列假設：

- (1) 具有有限的狀態空間 (finite state space)。
- (2) 穩定環境 (stationary environment)。
意即環境之行為與時間無關。雖然在現實環境下，穩定環境之性質並不常見，但本研究最主要之目的在驗證角色平等人機合作模式之合理性，並不打算處理環境變動之適應性問題。
- (3) 具有馬可夫 (Markovian) 性質之環境。由於本研究欲處理決策複雜、推理軌跡長之問題，在人腦屬於記憶有限 (short term memory)、可同時專注之資訊量有限的前提下，做此假設是合理的。
- (4) 無限區間 (infinite horizon) 推理過程。在推理過程尚未結束前，無法預先得知推理軌跡之長度。將利用報酬折扣率 (discount rate) 來處理無限推理區間下，預測期望報酬之問題。

根據上述四點假設性質，將本研究所面對的問題定義為馬可夫決策過程 (Markovian Decision Processes; MDPs) 問題。

欲達成本研究期望的成果，所設計之系統必須能處理符合前述之問題定義與本研究目的之相關議題。茲將本研究必須處理之議題與採行方式敘述如下：

- (1) 使用者參與其中，所以系統知識必須是使用者能理解者。本研究將採取網路表現法，以 feature base 之法

則方式來表現系統知識。並於系統運作過程中，透過使用者介面顯示系統推理所運用到的知識。

- (2) 系統必須有處理推理軌跡長問題的能力。透過網路表現法中，由“IF feature(s) THEN feature(s)”法則串成之連鎖推理 (chaining reasoning) 來處理。
- (3) 必須有動態性使用法則之能力 (dynamic rules)。系統在推理過程必須依據系統現處狀態，選用不同的法則知識。本研究利用每一個法則擁有之 Q 值 (Q-value)，根據 Q 值計算與修改每個法則之確定因子 (certainty factor)，逐步建立正確知識。推理過程中便依據確定因子高低，以決定性(deterministic)或機率性(stochastic)的方式，選取運用之法則。
- (4) 必須做到角色平等下使用者與系統溝通協調，相互學習。本研究透過設定法則門檻值 (rule threshold) 的做法，使系統根據選用法則之確定因子是否高於法則門檻值，來決定是否與使用者溝通。系統若對所用知識不確定，便向使用者學習；反之便堅持所選決策，不接受使用者介入教導。
- (5) 必須有處理報酬分配於不同時點的問題之能力。推理結束採取動作後，系統才知道決策之優劣，此時也才有回饋值 (reward)。由於採取 Q-learning 做為系統推理學習之基礎，欲加速學習就必須有效處理報酬分配於不同時點的問題。本研究利用使用者介入的過程，協助系統確認推理過程中不確定的步驟，重點修改這些不確定法則之確定因子，以使用者知識來有效處理報酬分配問題。

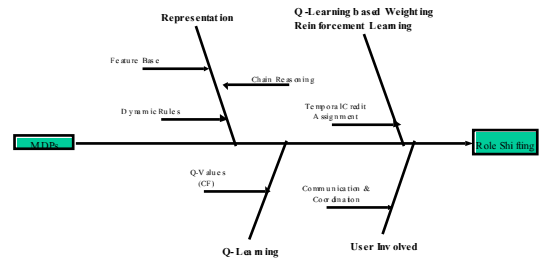


圖 1 問題解析魚骨圖

2.1 表現法的設計

本研究藉由法則方式來表現知識，並使每個法則擁有確定因子 (certainty factor; CF)。根據使用者介入之影響，來修改法則之確定因子，使得法則之使用具有動態性，以達到使用者與系統之主導角色可動態性移轉 (shifting roles) 之效果。本研究採取網路表現法作為主要的表現架構。

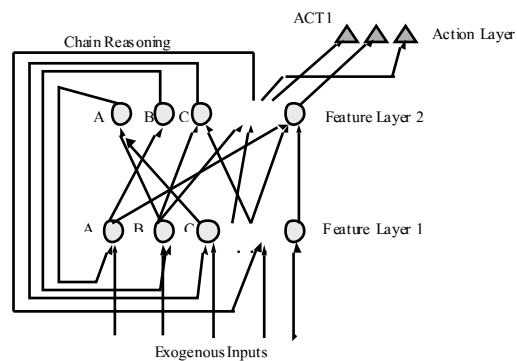


圖 2、網路表現法示意圖

可用一個節點代表一個特徵 (feature)，透過節點與節點間的連結來表示一個法則，利用連結權數 (connection weight) 來表現此法則之確定因子。將包含兩層的特徵階層 (Feature Layers)，此兩層所擁有的節點是完全相同的，Feature Layer1 代表法則「若」的部分 (IF part)，Feature Layer2 代表法則「則」的部分 (THEN part)；另外，還有一層的动作階層 (Action Layer)。其中，特徵代表系統在推理過程中的狀態；動作 (actions) 代表系統推理結束時所決定採取的最終決策。在此網路表現法下，將可以處理下列推理上法則運用問題：

1. Simple Feature/Action pair:
IF A THEN ACT1

2. Simple Feature/Feature pair: IF A THEN B

當系統開始推理過程，只要所運用到的法則之 THEN part 符合 Feature Layer2 中的節點時，則系統繼續連鎖推理下去。直到運用到的法則之 THEN part 符合 Action Layer 中的節點，推理過程方結束。只有 Feature Layer2 (THEN part) 中某些節點與 Action Layer 中的節點有連結。

2.2 決策主導地位動態移轉之設計

為了達到決策主導地位能動態性移轉，根據前述之系統設計原則，將決策主導地位動態性移轉之設計理念表示如圖 3。

當根據環境狀態開始推理時，先從以網路表現法表現之法則庫中比對符合環境狀態之法則。找到符合之法則後，根據這些法則以 Q-Learning 計算得之 Q 值，透過正規化的方式計算這些法則的 CF 值，並根據 CF 值選擇採用之法則。選擇運用之法則後，將此法則之 CF 值與法則門檻值比較，若高於法則門檻值便不詢問使用者建議，此時決策主導者為系統；若低於法則門檻值便詢問使用者建議，此時決策主導者為使用者。之後透過衝突化解機制運作，繼續推理過程。並於推理結束後根據環境回饋修改法則之 Q 值。

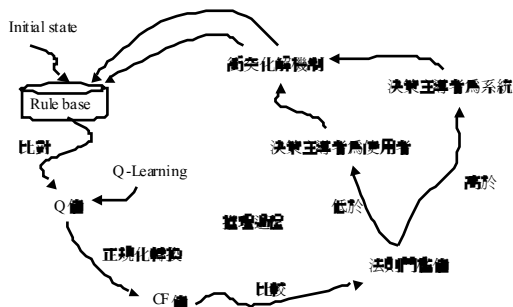


圖 3、決策主導地位動態移轉之示意圖

決策主導地位便隨著系統不斷根據學習結果修改法則 Q 值(CF 值)，以及每次推理時根據所選法則之 CF 值，與

法則門檻值比較之過程中，達到在使用者與系統間動態移轉之效果。

2.3 系統架構與運作方式

根據前述之問題定義、系統設計原則與表現法等，本研究所設計之角色平等人機合作模式系統架構如下圖：

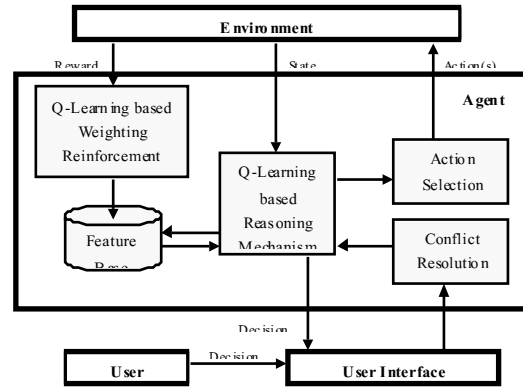


圖 4、角色平等人機合作模式架構圖

系統的運作流程如圖 5。

3、實驗設計及結果

為了驗證角色平等人機合作模式設計的效果，本研究採用一尋找收益最大路徑問題[11]做為實際驗證的測試平台。路徑尋找問題(Path Finding Problem)是作業研究領域中，時常被用來解說動態規劃法的問題。此問題的特性為必須在許多節點所組成的網路中，尋找從啟始節點至終止節點所組成的所有可能路徑中，成本最低或收益最大的路徑。

本研究以上圖有 10 個節點的路徑尋找問題為測試平台。從節點 1 開始，節點 10 結束，

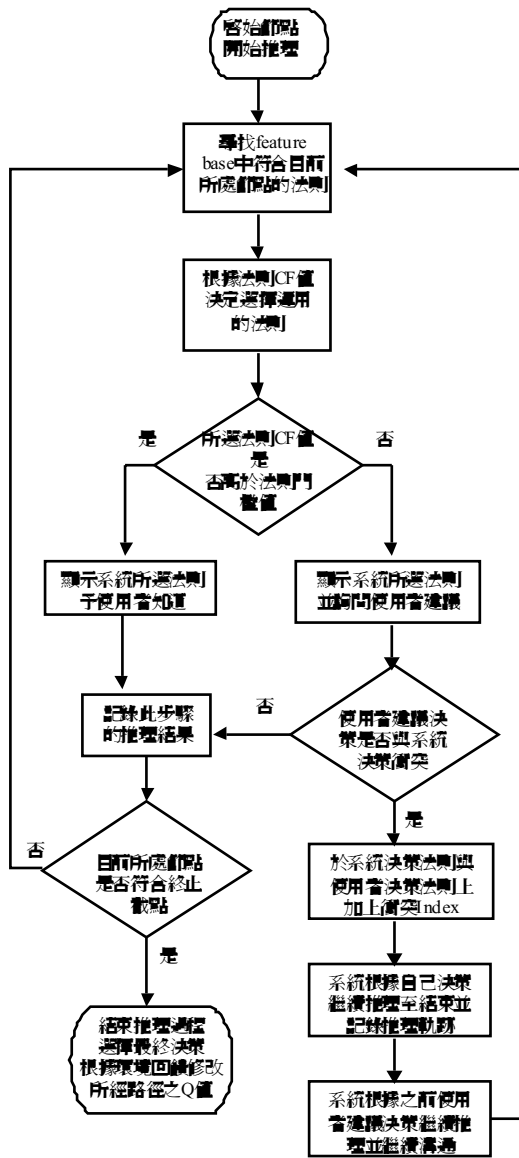


圖 5、系統運作流程圖

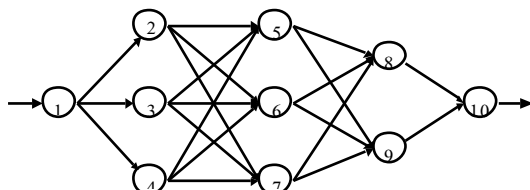


圖 6、路徑尋找問題示意圖

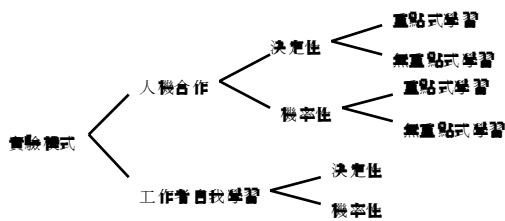


圖 7、實驗模式示意圖

從 18 條可能路徑中尋找收益最大之路徑。此路徑尋找問題從起點至終點共

有 9 個決策點(節點 1~9), 20 個可能決策(法則)。實驗共分成十組資料, 這十組資料共經由六種不同模式加以實驗測試。如圖 7 所示:

為了實驗結果的客觀性, 這十組資料於人機合作模式下的四種模式中, 共進行四十組實驗, 皆由相同的使用者參與系統運作。由於所採取的路徑尋找問題並非十分複雜, 因此在每一種模式下, 這十組資料皆以不同的順序加以測試, 使用者並無法得知共有幾組資料與目前為第幾組資料的資訊。在實驗執行次數方面, 工作者自我學習模式不分決定性或機率性, 每組資料每次實驗皆執行 50 次。人機合作模式方面, 決定性模式每組資料每次實驗皆執行 20 次, 機率性模式每次實驗皆執行 30 次。

根據本實驗的目的, 必須觀察系統在透過使用者的介入後, 是否有加速學習的現象; 此外, 必須觀察使用者與系統的決策是否漸趨一致, 且決策主導者是否有移轉的現象。我們首先觀察在六種不同模式實驗下, 第一次找到最佳解的執行次數, 以比較有無使用者介入對系統學習績效之影響。這裡所定義的第一次找到最佳解, 意義為不僅所找到的路徑為最佳路徑, 且決定此最佳路徑過程中所使用之法則其 CF 值皆高於法則門檻值。實驗的結果如表 1 所示。

接著, 我們將觀察工作者與使用者衝突和溝通次數之紀錄; 以及每次執行時, 最佳路徑決定者之紀錄。檢視是否確實產生決策漸趨一致與決策主導地位移轉之現象。

表 1、第一次找到最佳解執行次數比較表

實驗	Deterministic			Stochastic		
	無重點式學習	重點式學習	工作者自我學習	無重點式學習	重點式學習	工作者自我學習
1	16	11	21	23	9	41
2	12	9	***	16	8	22
3	**	9	***	29	13	***
4	9	7	***	14	7	24
5	10	9	***	21	7	17
6	13	8	***	10	14	24
7	14	8	***	9	8	20
8	16	10	17	12	7	20
9	11	5	***	14	8	27
10	17	13	***	12	9	22
平均	13.11	8.89		14.56	8.56	24.11

在決定性模式下，工作者除了第一組與第八組實驗可找到最佳解外，其餘各組皆掉入了局部最佳解。接下來觀察在機率性模式下工作者之表現，從表 1 可看出，由於機率性的導入，可協助工作者跳出決定性模式下之局部最佳解。在決定性模式下加入了使用者後，不論是有無重點式學習，表現皆比工作者自我學習來的好，皆能使工作者跳出局部最佳解的困境。

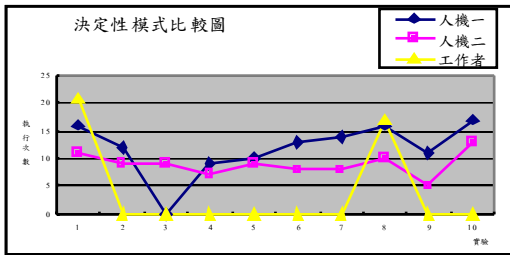


圖 8、決定性模式下有無使用者介入之比較註：人機一代表無重點式學習，人機二代表重點式學習。

若只比較重點式學習與無重點式學習兩種模式之表現，可發現重點式學習模式皆能以較少的次數便找到最佳路徑，如表 1 所示。從表中的平均次數可發現有重點式學習平均比無重點式學習快 4.22 次可找到最佳路徑。平均欄內資料為各模式去除實驗 3 資料之平均值。接著觀察機率性模式下有無使用者介入之比較。如圖 9 所示，在加入了使用者後，不論是有無重點式學習，除了無重點式學習模式在第五組實驗之表現稍遜於工作者自我學習外，其餘各組表現皆比工作者自我學習來的好，皆能以較少的次數找到最佳路徑。

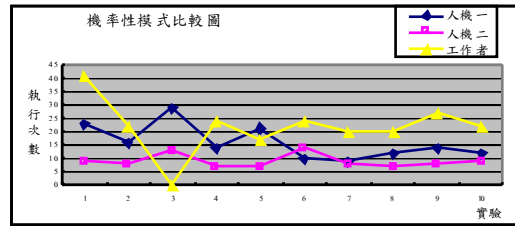


圖 9、機率性模式下有無使用者介入之比較

從表 1 的平均次數可發現 Stochastic 中有重點式學習平均比 Stochastic 中無重點式學習快 6 次可找到最佳路徑。可能原因為無重點式學習在 Q 值修改的過程中，受到隨機運作的影響較大，較不易收斂。而重點式學習因為在某些衝突點的 Q 值修改幅度較大，使得 Q 值收斂較快，CF 值之差異便隨之較快便大，隨機運作之影響便相對降低。

接著，我們比較重點式學習與無重點式學習分別在決定性模式下與機率性模式下之表現。從表 2 與表 3 中可看出來，不論是重點式學習模式或無重點式學習模式，在決定性模式下與機率性模式下表現皆相當接近。其平均找到最佳路徑之次數差距相當小。而從圖 10 與圖 11 中可看出，雖然不是十分明顯，但機率性模式找到最佳路徑之次數變動較大。

表 2、無重點式學習不同模式比較表

無重點式學習模式		
實驗	Deterministic	Stochastic
1	16	23
2	12	16
3	***	29
4	9	14
5	10	21
6	13	10
7	14	9
8	16	12
9	11	14
10	17	12
平均	13.11	14.56

表 3、重點式學習不同模式比較表

重點式學習模式		
實驗	Deterministic	Stochastic
1	11	9
2	9	8
3	9	13
4	7	7
5	9	7
6	8	14
7	8	8
8	10	7
9	5	8
10	13	9
平均	8.89	8.56

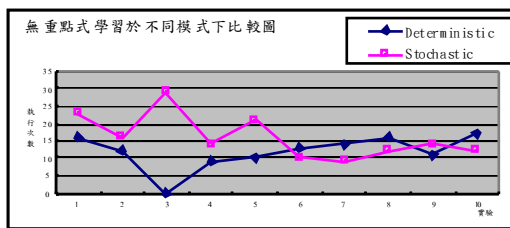


圖 10、無重點式學習於決定性與機率性下表現之比較圖

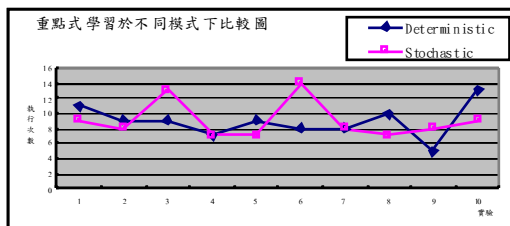


圖 11、重點式學習於決定性與機率性下表現之比較圖

雖然從實驗結果中可看出，重點式學習確實如期望中表現的比無重點式學習來的好，但是重點式學習在決定性模式下亦可能帶來落入局部最佳解的可能性。若使用者執意於某局部最佳解，在其一再介入的過程中，加重修改的方式，可能導引系統落入局部最佳解而無法跳出。

為了觀察是否確實有決策主導地位移轉以及使用者與工作者決策漸趨一致之現象，我們從無重點式學習模式中，挑出第一組。再從重點式學習模式中，挑出決定性模式之第二組之資料，以及機率性模式之第四組資料，加以觀察分析。

先看第一組的資料，如圖 12 與圖 13 所示。

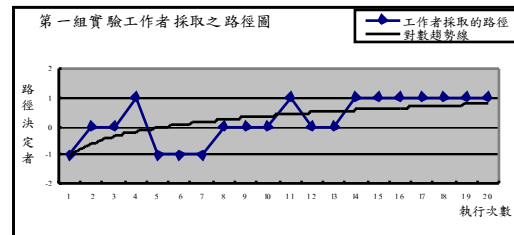


圖 12、第一組實驗工作者採取之路徑之軌跡圖

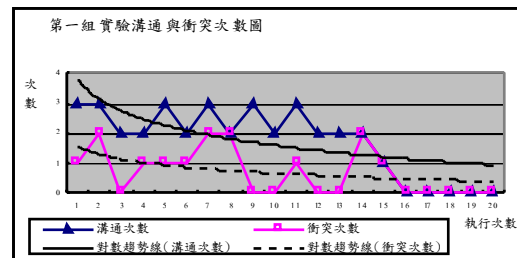


圖 13、第一組實驗工作者與使用者溝通及衝突次數圖

圖 12 中可看出，決策主導地位確實有漸漸從使用者移轉至工作者之現象。搭配圖 13 來看，圖中在第 4 次到第 12 次的執行次數中，工作者提出決策與使用者溝通的次數在 2 次與 3 次間震盪，之後便一路下降。對照至圖 12，在第 4 次到第 12 次的執行次數中，路徑決定者明顯地從使用者到雙方合作到工作者之順序移轉，之後便皆由工

作者決定而不再移轉。可說明工作者在第 4 次到第 12 次的執行次數中，Q 值正在大幅震盪，為其學習過程中最劇烈的時期。當第 13 次執行結束後，Q 值便逐漸收斂，到第 15 次執行結束，Q 值的收斂導致 CF 值一舉超過法則門檻值，系統便不再詢問使用者而相當堅持自己的決策。相同的情形可見於重點式學習模式（圖 14，圖 15），以及機率性模式（圖 16，圖 17）

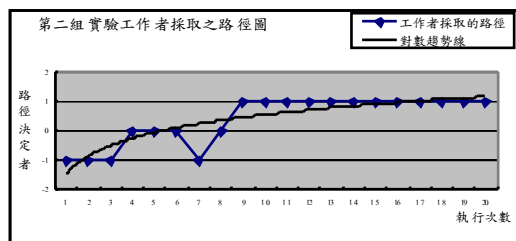


圖 14、第二組實驗工作者採取之路徑之軌跡

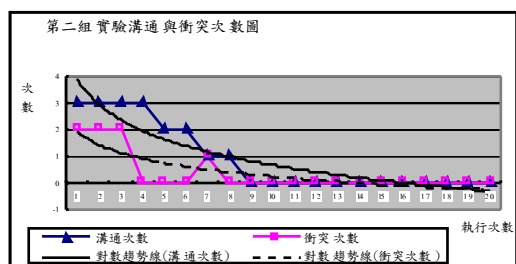


圖 15、第二組實驗工作者與使用者溝通及衝突次數圖

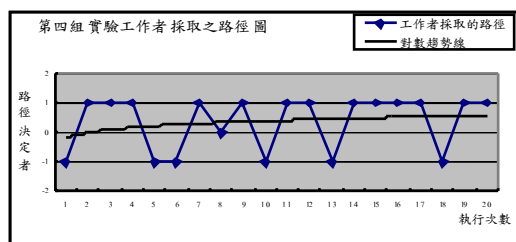


圖 16、第四組實驗工作者採取之路徑之軌跡

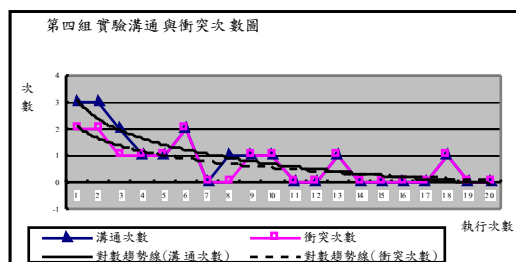


圖 17、第四組實驗工作者與使用者溝通及衝突次數圖

從實驗結果來看，確實產生了如本研究預期看到的現象。從第一次找到最佳解的次數來比較，不論是決定性模式或機率性模式，透過使用者的參與，確實能有效結合使用者知識，使工作者能在更少的次數內找到最佳解。同時，有使用者介入後，工作者有更大的機會跳出局部最佳解的困境。這是因為使用者的參與便扮演了機率性的角色。當使用者依據直覺作決策時，就如同機率性模式的運作。但使用者的直覺中，隱含了從系統運作過程所獲得訊息之判斷，亦即其直覺中隱含了對系統訊息之推理，故其直覺是屬於有方向性之直覺(或視為具有猜測正確方向之能力)。因此在使用者深耕策略與探勘策略的交互運用下，確實能導引工作者更快尋得最佳路徑。

4、結論與建議

本研究以路徑尋找問題做為所提出架構與機制之驗證，根據實驗結果，有下列結論：

1. 經由使用者介入，透過衝突化解機制運作，導引工作者找出推理過程不確定處，並予以加重修改，確實能使工作者快速學習到使用者有用的知識，減少工作者嘗試錯誤上所花費的時間，加速收斂。
2. 在衝突化解機制的作用下，工作者與使用者確實能透過溝通而相互學習，使決策漸趨一致。並能隨著決策信心的高低，動態性地移轉決策主導者地位。

經由實驗結果所獲得之結論，符合本研究所期望達到之目的。因此，本研究最大的貢獻，在於實際驗證了角色平等人機合作模式確實為合理可行之方式。同時，驗證了重點式增強學習方式確實為解決報酬分配於不同時點問題之可行方式。替未來對人機合作

模式有興趣之研究者，提供了一個參考的方向。

參考文獻

1. Decker, K. S. and V. R. Lesser, "Coordination Assistance for Mixed Human and Computation Agent System", *Umass Computer Science Technical Report 95-31*, 1995.
2. Dorigo, M. and H. Bersini, "A Comparison of Q-Learning and Classifier System", *Third International Conference on Simulation of Adaptive Behavior (SAB94)*, Brighton, UK, August 8-12, 1994.
3. Esquirol, P. and P. Lopez, "Constraint-oriented cooperative scheduling for aircraft manufacturing", *IEEE Expert*, pp. 32-39, 1997.
4. Kahn, G., "MORE: From observing knowledge engineers to automating knowledge acquisition", in *Automating Knowledge Acquisition for Expert Systems*, Chapter 2, Marcus, S.(ed), Kluwer Academic, 1988.
5. Masayuki Numao, "A Integrated Scheduling/Planning Environment for Petrochemical Production Processes", *Expert Systems With Applications*, Vol. 8, No. 2, pp.263-273, 1995.
6. Masayuki Numao, "Development of a Cooperative Scheduling System for the Steel-Making Process", *IEEE Expert*, pp.279-286, 1989.
7. Russell, S. J. and P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice-Hall, 1995.
8. Singh, S. P., "Learning to Solve Markovian Decision Processes", Ph. D. dissertation, Department of Computer Science University of Massachusetts, 1994.
9. Sutton, R. S. and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT, 1998.
10. Watkins, C. J.C.H. and P. Dayan, "Technical Note: Q-Learning", *Machine Learning*, 8, pp. 279-292, 1992.
11. Winston, W. L., *Operations Research: Applications and Algorithms*, Third Ed., Duxbury, 1993.
12. 吳信儀, 「以改良之進化策略演算法解決排序問題之研究—SRS 演算法與多重工作者系統之發展」, 東海大學工業工程研究所碩士論文, 民國八十五年.
13. 蔡佩芳, 「人機合作模式下的排程系統之研究—以分類者系統為架構」, 東海大學工業工程研究所碩士論文, 民國八十六年.

The Conflict Resolution Mechanism in Man-Machine Environment

Wei-Hua Andrew Wang
Department of IE, Tunghai University

Abstract

The most research in man-machine system are designed in giving the supreme decision privilege to the human (operator) or the machine. Since the status of the problem is varied, in this research, we have developed a man-machine-balanced cooperated design. The supreme decision privilege is shifted between man and machine dynamically according to the performance of the man-machine system as a whole.

Key words: Man-Machine, network-based rule representation, Q-learning, certainty factor

