

整合演化式計算與網際網路以支援軟體再用之研究

A Study of Integrating Evolutionary Computation with Internet for Software Reuse

計畫編號：NSC 87-2213-E-002-007

執行期限：86 年 8 月 1 日至 87 年 7 月 31 日

主持人：高成炎 國立台灣大學資訊工程學系

一、中文摘要

軟體再用系統必須包含搜尋組件、了解組件、修改組件和組合組件等四個功能，本研究建構一雛形以整合演化式計算與網際網路的技術以自動搜尋組件，以降低建立再用軟體資料庫的建構成本。我們建立的雛形是以整合智慧型助理、知識庫與資訊搜尋等技術，並採用演化式的機器學習技術為基礎，以結合統計和人工智慧技術的優點，試圖使本系統達到具動態彈性的優點，在本計劃中我們提出階層式的知識庫結構，此知識庫架構可分為共用分享層、群分享層、個人知識層與基本情況行為層等四層，我們以統計和演化式機器學習來更新和管理這些資料，根據實驗結果，此雛形系統可提升軟件搜尋精確度及達到軟件共享之目標。

關鍵詞：演化式計算、軟體再用、智慧型代理程式、資訊搜尋

英文摘要

A system of software reuse consists of four steps: retrieving components, modifying components, understanding components, and composing components. In this project, we construct a prototype in order to integrate evolutionary computation and internet to automatically retrieve reuse components and to reduce the cost of constructing a software reuse repository. The proposed prototype is

constructed by integrating intelligence agent, knowledge-base system, and information retrieval. We use an evolutionary approach in order to achieve the purposed of combining the statistical and artificial intelligence. Therefore, the proposed prototype is a dynamical and flexible system for software reuse. In this project, we propose a four-layer knowledge-based structure: shared layer, group layer, personal layer, and case behavior. At the same time, the statistic approach and evolutionary approaches are applied to update and manage the database. According to the experimental results, the proposed prototype can improve the recall and precision of software packages and achieve the purpose of sharing the software packages.

Keywords: Evolutionary Computation, Software reuse, Intelligence Agent, Information Retrieval

二、計畫緣由與目的

軟體再用是提升軟體生產力和品質的重要技術[10]，如能再使用軟體需求分析和設計的規格，可獲得更大的功效，物件導向具有封裝性、抽象性和模組化的特質，較容易達成軟體再用的目標[8]，且物件導向應用在系統分析、設計、撰寫程式的軟體生命週期具有容易整合的優點，而其結果自然具有明顯的階層結構，以系統

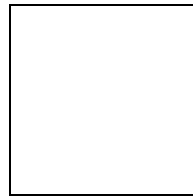
化的方法整合物件導向和軟體再用的技術，可提昇軟體生產力和品質。

演化式計算是以達爾文的「物競天擇」為基礎，主要可分為遺傳演算法[6]、演化式策略[1]和演化式規劃[5]等三種方法論，這些方法皆是健全且與問題獨立的機率演化式之最佳化方法論，因此他們可應用在多方面，包含組合性最佳化、函數係數的最佳化、工程控制和自動調適、機器學習等方面，而且大部份皆可求得令人滿意之近似最佳解的近似解。

網際網路於 1990 年後迅速發展，各種研究報告、學術論文、先進的研究觀念、各種程式庫和相關的資訊都相當豐富，而這些資訊大部份可透過網際網路取得，使得資訊分享的時代提早來臨，但如此句量的資訊，容易讓使用者迷失，如此反而造成使用者的不便，因此如何將這些加以歸類整理以配合使用者的需求，實為目前網際網路重要的研究主題，近幾年許多學者利用智慧型代理程式[3][4][9]試圖解決此問題，也獲得許多令人滿意的成就。本計劃將演化式計算[7][13]與資訊搜尋理論[2][7]加入智慧型代理程式，整合此三種技術以快速擷取再用組件，縮短建構軟體再用資料庫的時間，提升軟體生產力。

三、結果與討論

我們建立以智慧型助理、知識庫(knowledge base)與資訊搜尋(information retrieval)等技術，並採用演化式的機器學習技術演化式的機器學習技術為基礎代理程式的雛形，以結合統計和人工智慧的技術的優點，以避免純統計或純人工智慧的缺點[2]。此系統試圖達到容易使用及管理專家的知識。有許多研究顯示演化式計算是很容易和其他啟發式(heuristic)演算法結合，適當的啟發式演算法會明顯改進演化式計算的品質，尤其是對複雜的問題就更加明顯。演化式演算法可以提升智慧型助理和軟體再用的能力和彈性，而軟體再用技術可提升演化式演算法求解的品質，並增強智慧型助理的資料搜尋能力以滿足需求。



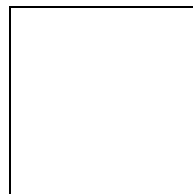
圖一：演化式運算發展環境之智慧型助理的雛形架構

為達成及提升演化式計算整合環境中各種知識與軟體的重複使用，在本計劃中我們提出階層式的知識庫結構，階層的數目依應用環境而定，基本的知識庫架構可分為共用分享(common share)層、群分享(group share)層、個人(personal)知識層與基本情況行為(case behavior)層等四層，我們以統計和演化式機器學習來更新和管理這些資料。如此可以自動提供適時且適量的資料，以滿足使用者提出的需求，同時可提高智慧型助理的彈性、調適、個別化與經驗分享等能力。

整個演化式演算法發展系統之智慧型助理的架構，如圖一所示，下面各節將分別說明如何結合物件導向、智慧型助理及知識庫、演化式動態的搜尋系統，以達成軟體再用的表目標。

A. 階層式軟體再用與物件導向的結合

本研究首先利用資訊搜尋 (I R) 的技術，自動將網際網路蒐集到的相關資料加以分析，建立以構面因子[11](Facet scheme)和關鍵字(keyword)為基礎的索引，並根據群聚理論 (H A C) [12]將再用組件依相似程度加以分類，其結果和演算法如圖二所示，以提升資料搜尋的效率與精確度，同時以演化式演算法的技術達成動態管理的目標。



1. Let each component be a cluster.
2. Find the pair of clusters that has the maximum value of similarity among all of clusters
3. Merge the pair of clusters found in the step into a new cluster

4. If there is only one cluster or $\leq$ lower limited cluster, then stop. Otherwise, Update similarity value between clusters; go back step 2

圖二：The HAC algorithm and constructing structures

B. 智慧型助理與知識庫的表示

智慧型助理重要的兩個特點就是學習能力和個人化的輔助[3][4]，因此需要適當的知識庫與機器學習的方法，以記錄使用者的習慣、興趣和特殊的習慣，同時也要記錄使用者共同需要的相關知識、共同的問題解答、使用說明等，以提供適時適量的資訊給使用者。機器學習的技術，使得智慧型助理能在使用者遇到問題時提供一些建議、保持最新的動態資訊以滿足需求、隨環境變化的調適能力。

我們以四層的知識庫架構，以支援智慧型助理的需求：

第一層：行為資料的儲存層：此層資料直接記錄使用者的使用行為，所有系統的使用行為，智慧型助理皆會以知識表示法加以儲存，在本系統會以簡單的（使用者，關鍵詞，選擇，行為序號）的方式加以儲存，智慧型助理可利用這些資料，利用統計和機器學習的技術，歸納使用者與使用群體的習慣和偏號，將某些知識提升至第二或第三層，以達到個人化與小組化的特殊輔助功能。

第二層：個人化的知識庫：每個人皆有特殊的習慣與興趣，因此需要這層的知識，以滿足個人化輔助的需求，這是智慧型助理直接提到的需求，尤其在個人電腦非常大眾化的時代，這種需求更是迫切。使用者可以預設某些知識或習慣，作為智慧型助理輔助的依據，使用者也可設定條件，作為智慧型助理特殊處理。當然智慧型助理會學習使用者的習慣，自動將這些知識放入知識庫內。知識庫的知識皆有強弱的加重比率，作為提供建議的衡量標準，以提高智慧型助理的輔助品質。

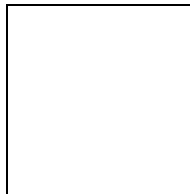
第三層：小組的知識庫：每個小組的人除需保留個人的任務外，整個小組也有共同的目標，因此需要這層的知識，以滿

足小組化輔助的需求。小組共同的行為和認知，形成小組的文化，因此有這層知識庫使得共同的知識可以分享，經驗得以傳承，同時可以縮短同組人員的學習時間。

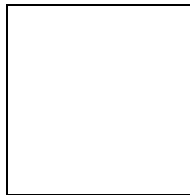
第四層：共享知識庫：此層的知識由專家及系統設計者提供，智慧型助理也會根據使用者的共同習慣，自動將這些知識放入知識庫內，知識庫的知識也有強弱的加重比率，作為提供建議的衡量標準，以提高智慧型助理的輔助品質。

C 演化式的動態搜尋和機器學習技術

動態索引和調整使用者需求的目的是要更精確找出使用者的需求如圖三所示，而我們建議以演化式和知識庫的技術來達成這個目標，因此知識庫和軟體再用組件皆以關鍵字作為搜尋的條件，且每個關鍵字的比重依重要性分別定義不同的比值，以配合演化式技術(FCEA)的應用。此FCEA是我們所提出的演化式技術，已成功用應用在整體最佳化[14]及類神經網路[15]，其架構如圖四所示。

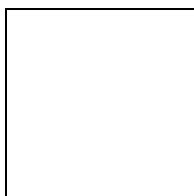


圖三：動態索引與重新組合查詢的示意圖



圖四：FCEA演化式演算法示意圖

染色體的表示方式以每個知識或再用組件，皆以構面因子加以描述，關鍵字的比值 $W(0 \leq W \leq 1)$ ，故以實系數來表示較為直接。整個系統的查詢過程和資料流程如圖五所示，粗體部份為演化式的機器學習部分和智慧型助理的核心。



圖五：演化式動態搜尋和機器學習的示意圖

四、參考文獻

1. T. Bäck, *Evolutionary Algorithms in Theory and Practices*, Oxford, 1995.
2. W. B. Croft 1993, "Knowledge-base and statistical approaches to text retrieval," *IEEE Expert*, pp. 8-11, Apr. 1993.
3. O. Etzioni and D. Weld, "A Softbot-Based Interface to the Internet," *CACM*, vol. 37, no.7, pp72-76. July,1994.
4. O. Etzioni, and D. Weld, "Intelligent Agents on the Internet: Fact, Fiction, and Forecast," *IEEE Expert*, pp44-49. Aug.m 1995.
5. D. B. Fogel, *Evolutionary Computation: Toward a new philosophy of machine intelligence*, IEEE Press,1995.
6. D. E. Goldberg, *Genetic Algorithms in search, Optimization & Machine Learning*, Reading. MA: Addison-Welsley,1989.
7. M. Gordon, "Probabilistic and Genetic Algorithms for Document Retrieval," *CACM*, vol.3, no. 10, pp. 1208-1218, Oct. 1988.
8. I. Jacobson, *Object-Oriented Software Engineering - A use Case Driven approach*, Addison-Wesley, Wokingham, 1992.
9. P.Maes, "Agents that Reduce works and Information Overload," *CACM*, vol. 37, no.7, pp31-40, July, 1994.
10. B. Meyer, "Reusability:The case for object-oriented design," *IEEE Software*, vol. 4,no. 2, pp. 38-51, Mar. 1987.
11. R. Prieto-Diaz and P. Freeman, "Classifying software for reusability," *IEEE Software*, vol. 4, no. 1, pp. 6-16, Jan. 1987.
12. G.Salton and M.J.McGill, *Introduction to Modern Information Retrieval*. New York: McGraw-Hill,1983.
13. J. J. Yang and R. R. Korfhge "Query Optimization in Information Retrieval Using genetic Algorithms," *ICGA* pp. 603-611, 1993.
14. J. M. Yang, C. Y. Kao and J. T. Horng, "A continuous genetic algorithm for global optimization." In *Proceeding of the Seventh Intl. Conference on Genetic Algorithm* 1997, 230-237.
15. J. M. Yang, J. T. Horng, and C.Y. Kao, "A New Evolutionary Approach to Developing Neural Autonomous Agents," to appear In *1998 IEEE Intl. Conference on Robotics and Automation*, 1998.