國立交通大學

資訊科學系

碩 士 論 文

一個針對二維資料的機率式重新分配模擬退火分群法的介面

A GUI of Simulated-Annealing K-means Clustering with Probabilistic Redistribution for 2D Data

研 究 生：卓明宏

指導教授：李嘉晃　教授

中 華 民 國 九 十 三 年 七 月

一個針對二維資料的機率式重新分配模擬退火分群法的介面

A GUI of Simulated-Annealing K-means Clustering with Probabilistic Redistribution for 2D Data

研 究 生：卓明宏　　Student：Ming-Hung Cho

指導教授：李嘉晃　　Advisor：Chia-Hoang Lee

國 立 交 通 大 學

資 訊 科 學 系

碩 士 論 文

A Thesis

Submitted to Institute of Computer and Information Science

College of Electrical Engineering and Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer and Information Science

July 2004

Hsinchu, Taiwan, Republic of China

中華民國九十三年七月

一個針對二維資料的機率式重新分配模擬退火分群法的介面

研究生 ： 卓明宏　　指導教授 ： 李嘉晃 教授

國立交通大學電機資訊學院　資訊科學系碩士班

中文摘要

分群法是一個被廣泛使用的自動化資料分類技術，常用於機器學習或資料探勘的先置處理步驟前置。其中模擬退火法是一種模擬多粒子物理系統由較高能量熱平衡狀態降溫到較低能量熱平衡狀態過程以取得最少 TSSE 的分群法。SAKM-clustering 整合了模擬退火法取得系統最小能量及 K-means 快速搜尋的能力。它適用於搜尋多維特徵空間中適當分群並使得分群結果在相似度測度上最佳化。在本論文中我們針對二維資料實作了一個界面以觀察 SAKM-clustering 及 K-means 的分群過程效能比較。

A GUI of Simulated-Annealing K-means Clustering with Probabilistic Redistribution for 2D Data

Student : Ming-Hung Cho Advisor : Prof. Chia-Hoang Lee

Department of Computer and Information Science

National Chiao Tung University

ABSTRACT

Clustering is an extensively used technique for automatic data classification, such as in the preprocessing of machine learning and data mining. Simulated-annealing is a clustering technique, which obtains the minimum TSSE of a group of data by simulating the cooling down process of a many-particle physical system from a state in thermal equilibrium with higher energy to another state in thermal equilibrium with lower energy. The SAKM-Clustering integrates the power of simulated-annealing for obtaining minimum energy configuration and the searching capability of K-means algorithm. It is used to search proper clusters in multidimensional feature space such that a similarity metric of the resulting clusters is optimized. In this thesis we implement a GUI to observe the clustering processes and to compare the performances of the SAKM-clustering as well as the K-means clustering.

誌謝

本論文得以順利完成，首先要感謝的是指導教授李嘉晃老師，他給予我不斷的機會去嘗試，給予我充分的空間發展，並且給予我最大的耐心與包容。要特別謝謝老師讓我能有機會輔修電子物理系的課程，完成多年來追尋基礎科學課程的願望。也感謝口試委員蘇豐文教授與楊武教授的指正建議，讓我獲益良多。

要謝謝國立交通大學的師長們及許多單位的長官與工作人員，他們提供了一個高度專業與安全舒適的環境讓我可以安心於課業，並且可以追尋自己所喜愛的知識。要謝謝資訊科學系及電子物理系的助理小姐們熱心的幫助與關心。要謝謝電子物理系授課老師們的教導與幫助，讓我能有機會一窺這個神秘殿堂的奧妙與神奇之處。

還要感謝實驗室所有學長，同學與學弟們的幫助，充實了我的實驗室生活。尤其要謝謝湯隆宜，林駿豪，王明德，高鳴遠，黃明超，林建良幾位，謝謝大家幫我度過許多忙亂的階段。一路走來曾幫助我的老師，助教，學長姊及交大清大的同學們，我也在此表達對各位誠摯的謝意。

最後我要感謝我的家人，因為他們多年來的支持與鼓勵是我努力的最大動力。謹以此論文獻給我最親愛的家人。

目錄

圖目錄