

1.1 | 研究目的

現代分散式資料庫減緩了單個服務器資料讀寫的壓力，如果將單個資料庫結構拆分成多個資料庫表並分散到各個資料庫集群，需要良好的資料關聯分離以達到高讀寫、低延遲、易擴展等效果。此研究使用演化演算法的聚類方式將資料進行關聯性離散化，並將關聯性高的資料進行資料庫表分類，最後利用演化演算法調度分散式資料庫的調度演算法。

1.2 | 研究背景

2012 年，維多利亞·帕雄·阿爾瓦雷斯提出了一種進化工具，用於在包含定量和分類屬性的數據庫（無論大小）中查找關聯規則，而無需對數字屬性的域進行先驗離散化。這是為了解決數據庫非常大的時候，很多關聯規則工具都無法使用 [1]，於是使用基於演化演算法理論的 GAR-plus 工具 [2] 找到最佳規則參數。

2020 年，Forhad Zaman 博士在大型專案的混合進化演算法調度問題中，提出了一種有效的混合算法，其中兩個多算子進化演算法在兩個亞群下順序形成，它們的大小根據它們的動態調整進化過程中的表現，此外，提出了兩種啟發式方法，第一種一種是基於線性規劃方法，旨在獲得可行的模式，而第二個是基於改進的前向和後向對齊方法，目的是獲得可行的時間表 [3]。

1.3 | 研究方法

先尋找開源的資料庫資料集合，提取資料特徵並資料數字離散化，先使用 k-means 聚類算法將其進行層層進化分類，在不斷的訓練模型期間，改善演算法的缺陷或未完善的參數。根據資料的訓練結果將其分佈在分散式資料庫集群，嘗試使用 MMRCPS 蟻群演算法規劃調度各個資料庫的備份調度策略。

1.4 | 總結

臺灣師範大學的師資陣容和教育資源吸引我前往進行繼續的深造，包括資料庫、演算法、雲計算等領域都有所建設。本次的研究計畫也是針對我未來在研究所的未來課題進行思考，而貴校的資源非常匹配本次的研究計畫，讓我非常的心動，也想要增強自己對於演算法與機器學習、人工智能方面的知識盲點，融入校園，實現人生理想。

1.5 | 參考文獻

- 1 Pachón Álvarez, V., Mata Vázquez, J. (2012). An evolutionary algorithm to discover quantitative association rules from huge databases without the need for an a priori discretization. *Expert Systems with Applications*, 39(1), 585–593. doi:10.1016/j.eswa.2011.07.049
- 2 Goldberg, E. D. (1989). *Genetic algorithms in search. Optimization and machine learning*. Wesley Longman.
- 3 Zaman, F., Elsayed, S., Sarker, R., Essam, D. (2020). Hybrid Evolutionary Algorithm for Large-scale Project Scheduling Problems. *Computers Industrial Engineering*, 106567. doi:10.1016/j.cie.2020.106567