

PA  
2200  
1002  
HG

# 並列処理による高度統計データ解析に関する研究

下平 文彦

1999年7月

寄贈  
下平文彦氏

00002525

# 目次

<b>1</b>	<b>序論</b>	<b>3</b>
1.1	近年のデータ解析	3
1.2	これまでの統計データ解析の問題点	3
1.3	本研究の目的	3
1.4	論文の構成	4
<b>2</b>	<b>統計データ解析および並列化プログラム</b>	<b>5</b>
2.1	統計データ解析手法の分類	5
2.2	基本統計量	6
2.2.1	基本統計量とは	6
2.2.2	基本統計量の並列化	6
2.2.3	相関係数を求める処理の拡張	8
2.3	重回帰分析	8
2.3.1	重回帰分析とは	8
2.3.2	重回帰分析の適合性の指標	9
2.3.3	重回帰分析の並列化	10
2.4	主成分分析	10
2.4.1	主成分分析とは	10
2.4.2	主成分分析モデルの定式化	12
2.4.3	寄与率・累積寄与率	13
2.4.4	主成分得点	14
2.4.5	主成分分析の並列化	14
2.5	判別分析	15
2.5.1	判別分析とは	15
2.5.2	線形判別関数を用いた判別分析	15
2.5.3	正準判別分析	15
2.5.4	判別分析の並列化	17
2.5.5	並列処理での利点	19
2.6	数量化理論 I 類	19
2.6.1	数量化理論 I 類とは	19
2.6.2	数量化理論 I 類の並列化	19
2.7	数量化理論 II 類	22
2.7.1	数量化理論 II 類とは	22
2.7.2	数量化理論 II 類の並列化	22
2.8	クラスタ分析	24
2.8.1	クラスタ分析とは	24

2.8.2	クラスタ分析の並列化	25
2.9	開発環境	26
2.9.1	MPI	26
2.9.2	LAM	26
<b>3</b>	<b>並列計算機システムの性能評価</b>	<b>27</b>
3.1	実行環境	27
3.1.1	CP-PACS	27
3.1.2	RS/6000 SP	28
3.1.3	パソコン・クラスタ	28
3.2	並列処理の評価基準	29
3.3	性能評価実験	30
3.3.1	測定に用いた解析	30
3.3.2	各環境での特性	31
3.3.3	CP-PACS,RS/6000 SP,パソコン・クラスタの相対性能評価	36
3.4	まとめ	45
<b>4</b>	<b>システム利用環境</b>	<b>46</b>
4.1	操作インタフェース	46
4.1.1	操作システム概要	46
4.1.2	開発環境	46
4.1.3	実行例	46
4.2	結果出力	48
<b>5</b>	<b>実データを用いた処理の例</b>	<b>52</b>
5.1	判別分析の例	52
5.2	相関係数を求める処理の例	52
5.2.1	区間によって, 正・負の相関があらわれる例	56
5.2.2	ある区間にだけ, 負の相関があらわれる例	58
5.3	重回帰分析を行った例	60
5.3.1	Adult literacy rate に対する解析結果	62
5.3.2	Real GDP per capita に対する解析結果	63
5.3.3	Education index に対する解析結果	65
5.3.4	GDP index に対する解析結果	67
5.3.5	Human development index (HDI) value に対する解析結果	69
5.3.6	GDP に関連する項目に対する解析結果	71
<b>6</b>	<b>結論</b>	<b>76</b>
	謝辞	78
	参考文献	79
	公表論文	82