

N° d'ordre : 61

THÈSE

présentée devant L'ÉCOLE NORMALE SUPÉRIEURE DE LYON *pour*

obtenir le Titre de Docteur de l'École Normale Supérieure de Lyon

**spécialité : Informatique au titre de la formation Doctorale
d'Informatique de Lyon par Didier PUZENAT**

Parallélisme et Modularité des Modèles Connexionnistes

Soutenue le 20 octobre 1997

Après avis de : ALEXANDRE Frédéric
 JUTTEN Christian

Devant la commission d'examen formée de :

ALEXANDRE	Frédéric
AMY	Bernard
COSNARD	Michel
JUTTEN	Christian
PAUGAM-MOISY	Hélène
ZIGHED	Djamel

Sommaire

Avant Propos	1
Remerciements	3
Résumé	5
Introduction	7
De bonnes raisons de s'intéresser au parallélisme	7
Un classifieur incrémental	7
Généricité des parallélisations	8
Les voies de la parallélisation	8
Une approche expérimentale	8
Sciences cognitives	9
Plan commenté	9
1 Classification et réseaux de neurones artificiels	11
Introduction	11
1.1 Classifieur neuronal incrémental	13
1.1.1 Réseaux incrémentaux et littérature connexionniste	14
1.1.2 Architecture du classifieur incrémental	14
1.1.3 Variante de l'algorithme incrémental	17
1.2 Évaluation du classifieur en reconnaissance de formes	18
1.2.1 Un prétraitement adapté aux données	18
1.2.2 Algorithme d'apprentissage supervisé adapté à l'OCR	19
1.2.3 Performances du classifieur	20
1.2.4 Conclusion sur le problème d'OCR	22
1.3 Évaluation du classifieur sur les « formes d'ondes »	22
1.3.1 Construction des formes d'ondes	22
1.3.2 Algorithmique du classifieur pour les formes d'ondes	23
1.3.3 Performances du classifieur	24
1.3.4 Conclusion sur le problème des formes d'ondes	26
Conclusion	27

2	Parallélisme et informatique distribuée	29
	Introduction	29
2.1	Parallélisme pour le modélisateur	30
2.2	Parallélisme pour l'ingénieur connexionniste	30
2.3	Pour aller plus vite	31
	2.3.1 Circuit de neurones artificiels VLSI	31
	2.3.2 La « concurrence » du séquentiel	32
2.4	Choix d'une architecture parallèle	33
2.5	La Volvox IS-860	34
2.6	Machines parallèles contre stations de travail	37
2.7	Le système PVM	37
2.8	PVM en perspective	40
	Conclusion	42
3	Parallélisation du classifieur par partage de l'espace d'entrée	45
	Introduction	45
3.1	Granularité du partage de l'espace d'entrée	46
3.2	Le prétraitement du problème d'OCR	47
	3.2.1 Détail du prétraitement	47
	3.2.2 Parallélisation du prétraitement	49
3.3	Activation des prototypes en parallèle	51
	3.3.1 Partage des prototypes	51
	3.3.2 Distribution des prototypes	53
	3.3.3 Choix d'une stratégie	54
3.4	Mise en œuvre du classifieur à granularité moyenne	55
3.5	Étude du temps de prétraitement en apprentissage	57
	Conclusion	58
4	Parallélisation modulaire du classifieur incrémental	61
	Introduction	61
4.1	Approche modulaire	61
4.2	Parallélisation de la généralisation	63
4.3	Indépendance des prototypes	63
4.4	Parallélisation de l'apprentissage	65
4.5	Comportement du classifieur durant l'apprentissage	66
4.6	Accélération atteinte	68
4.7	Amélioration de l'accélération en apprentissage	69
4.8	Solution dans le cas d'un apprentissage supervisé	70
4.9	Équilibrage des calculs	73
	Conclusion	77

5	Parallélisation du classifieur sur une machine parallèle virtuelle	81
	Introduction	81
	5.1 Parallélisation modulaire sur réseau de stations	81
	5.2 Disponibilité des exemples	83
	5.3 Algorithme d'apprentissage parallèle asynchrone	85
	5.4 Dangers d'un apprentissage spécialisé asynchrone	86
	5.5 Tolérance aux pannes	89
	5.6 Apprentissage asynchrone des formes d'ondes	91
	5.7 Accélération sur le problème des formes d'ondes	92
	5.8 Analyse des ensembles d'apprentissage	97
	Conclusion	100
6	Étude du traitement cognitif des odeurs	101
	Introduction	101
	6.1 Expérimentations auprès des sujets	102
	6.2 Utilisation du classifieur	104
	6.3 Résultats des tentatives de classification	105
	6.4 Analyse des données olfactives	107
	6.5 Distances et coefficients de corrélation-croisée	108
	6.6 Un nouvel algorithme d'apprentissage incrémental	110
	6.7 Mise en œuvre d'un réseau d'ondelettes	111
	Conclusion	113
7	Modularité et processus cognitifs de mémorisation	115
	Introduction	115
	7.1 Un modèle des processus de mémorisation humains	115
	7.2 Simulation d'un amorçage de répétition	117
	7.2.1 Hypothèse de la confiance	118
	7.2.2 Détection et mesure de l'amorçage	119
	7.2.3 Expérience d'amorçage du classifieur	120
	7.2.4 Étude de la fonction d'amorçage	123
	7.2.5 Conclusion sur l'amorçage	124
	7.3 Perspective d'une mémoire associative hétérogène	125
	7.3.1 Module perceptif	126
	7.3.2 Module associatif	128
	7.3.3 Processus d'évocation	129
	Conclusion sur une mémoire associative hétérogène	130
	Conclusion et perspectives	133
	Parallélisme et connexionnisme	133
	Parallélisme et classification incrémentale	133
	Apports mutuels des sciences cognitives et du connexionnisme	135

Mémoire associative hétérogène	135
A Bases du connexionnisme	137
Introduction	137
A.1 Le neurone formel de McCulloch et Pitts	137
A.2 Le perceptron de Rosenblatt	138
A.3 Du neurone au réseau de neurones	139
A.4 La nécessité d'un apprentissage	142
A.5 Notion de prototype	143
A.6 Apprentissage par compétition	144
A.6.1 Les cartes de Kohonen	144
A.6.2 Apprentissage LVQ	145
B Bases du parallélisme	147
Introduction	147
B.1 Les machines parallèles selon Flynn	147
B.2 Quelques définitions	148
B.2.1 Capacité d'évolution (« scalabilité »)	148
B.2.2 Accélération (<i>Speedup</i>)	148
B.2.3 Latence et bande passante	149
B.2.4 Portabilité	150
B.2.5 Mise en œuvre d'un pipeline	150
B.2.6 Granularité et parallélisme	151
Diffusion des travaux réalisés	153
Bibliographie	155

Table des figures

1.1	Architecture du classifieur neuronal incrémental	15
1.2	Schéma fonctionnel du classifieur incrémental	16
1.3	Exemples de chiffres manuscrits à apprendre et à reconnaître	18
1.4	Couplage du module de prétraitement et du classifieur incrémental	19
1.5	Évolution des temps de traitement en apprentissage	21
1.6	Trois formes d'ondes dont la combinaison engendre trois classes	23
2.1	Architecture parallèle MIMD	33
2.2	Un nœud de la Volvox est composé de deux processeurs	35
2.3	Un ordinateur central puissant et des terminaux	40
2.4	Exemple d'un réseau de stations (avec un terminal)	41
2.5	Principe du partage du réseau	43
2.6	Principe de la distribution des exemples	43
3.1	Configuration de 8 processeurs en anneau sur une Volvox	46
3.2	Douze groupes de cellules détectent douze orientations	47
3.3	Exemple d'image discrétisée (pixels) en entrée du système	47
3.4	Activation des cellules à champ récepteur des 12 orientations	47
3.5	Chaque cellule à champ récepteur scrute une partie de l'image	48
3.6	Influence des connexions latérales sur les segments détectés	49
3.7	Sortie du module de prétraitement (entrée du classifieur)	49
3.8	Sous-images confiées aux processeurs	50
3.9	Diffusion des segments orientés avec un masque $3 \times 3 \times 3$	50
3.10	Activation des segments orientés avec un masque 5×5	51
3.11	Chaque processeur possède tous les prototypes	52
3.12	Chaque processeur possède des prototypes différents	53
3.13	Multidiffusion des blocs de cellules à champ récepteur	56
3.14	Temps de prétraitement et de classification avec 2 processeurs	57
3.15	Temps de prétraitement et de classification avec 4 processeurs	57
3.16	Schéma complet de la parallélisation par partage de l'espace d'entrée	60
4.1	Chaque processeur reçoit une partie de l'ensemble des N exemples	62
4.2	Pourcentage d'exemples reconnus en fonction du nombre de modules	66

4.3	Nombre de prototypes créés en fonction du nombre de modules	67
4.4	Accélération en apprentissage et en généralisation	68
4.5	Temps de prétraitement, de classification, et de communication	69
4.6	Temps d'apprentissage en fonction du délai (δ)	72
4.7	Accélérations en apprentissage en fonction du délai (δ)	72
4.8	Pourcentage d'exemples reconnus en fonction du délai (δ)	73
4.9	Temps cumulés de calcul et communication en fonction de δ	76
4.10	Schéma complet de la parallélisation modulaire	80
5.1	Réseau local de stations de travail (bus)	83
5.2	Parallélisation de l'acquisition et libération du réseau local	84
5.3	Utilisation d'une information locale pour contrôler l'avance des modules	88
5.4	Nombre de prototypes créés localement, avec $\delta = 1$ et $\lambda = 1$	93
5.5	Nombre total de prototypes créés avec $\delta = 1$ et $\lambda = 1$	93
5.6	Nombre de prototypes créés localement, avec $\delta = 20$ et $\lambda = 1$	94
5.7	Nombre total de prototypes créés avec $\delta = 20$ et $\lambda = 1$	94
5.8	Nombre de prototypes créés localement, avec $\delta = 1$ et $\lambda = 20$	95
5.9	Nombre total de prototypes créés avec $\delta = 1$ et $\lambda = 20$	95
5.10	Nombre de prototypes créés localement, avec $\delta = 20$ et $\lambda = 20$	95
5.11	Nombre total de prototypes créés avec $\delta = 20$ et $\lambda = 20$	95
5.12	Pourcentage d'erreurs en généralisation, en fonction de δ et de λ	96
5.13	Erreur, en fonction de δ	96
5.14	Erreur, en fonction de λ	96
5.15	Deux situations caractéristiques ne nécessitant pas de communications	97
5.16	Sans communications	99
5.17	Avec communications	99
6.1	Exemples de choix de sujets ayant bien répondu	107
6.2	Exemples de choix de sujets distincts ayant mal répondu	107
6.3	Quatre exemples très similaires pour quatre odeurs de fruit distinctes	108
6.4	Deux exemples assez similaires pour deux odeurs très différentes	108
6.5	Mesures de distances pour les différents exemples d'une odeur donnée	109
6.6	Coefficients de corrélation-croisée de l'orange	110
6.7	Architecture d'un réseau d'ondelettes	112
6.8	Performances obtenues par réseau d'ondelettes	113
7.1	Schéma fonctionnel, simplifié à l'extrême, de la mémoire humaine	116
7.2	Passages dans le module perceptif	117
7.3	Passage dans deux modules perceptifs différents	118
7.4	Fonction d'amorçage f	119
7.5	Fonction f discrétisée	119
7.6	Évolution de la confiance en fonction des délais réalisés	121

7.7	Confiance moyenne avec et sans amorçage	121
7.8	Occurrence des délais	121
7.9	Pourcentage d'occurrence de chaque délai	121
7.10	Confiance moyenne sur 150 cycles apprentissage-généralisation	122
7.11	Pourcentage d'apparition des délais	122
7.12	Évolution de la pente de la droite de régression	123
7.13	Différentes bases pour la fonction d'amorçage f	123
7.14	Pentes de régressions pour les différentes bases	123
7.15	Différentes hauteurs pour la fonction d'amorçage f	124
7.16	Pentes de régressions pour les différentes hauteurs	124
7.17	Architecture minimale compatible avec le modèle	126
7.18	Une mémoire à deux niveaux	127
A.1	Un neurone formel	138
A.2	Fonction ET	139
A.3	Fonction OU	139
A.4	Fonction OU <i>exclusif</i>	139
A.5	Exemple de réseau <i>feedforward</i> à couches	140
A.6	Exemple d'architecture pour un réseau récurrent	141
A.7	La fonction sigmoïde	143
A.8	Un prototype et la classe qui lui est associée	143
A.9	Hyperplans séparateurs	144
A.10	Hypersphères d'influence	144
B.1	Architecture séquentielle SISD	147

Liste des tableaux

1.1	Qualité de l'apprentissage séquentiel sur la base de chiffres	21
1.2	Temps d'exécution de l'apprentissage séquentiel sur la base de chiffres	21
1.3	Temps de prétraitement et de classification en fin d'apprentissage . .	22
1.4	Résultats sur les formes d'ondes (base originelle)	24
1.5	Effectif de chacune des trois classes du problème de Breiman <i>et al.</i> .	24
1.6	Résultats sur les formes d'ondes	25
1.7	Comparaison du classifieur avec les modèles du projet inter-PRC . .	26
2.1	Latence et taux de transfert de la Volvox	36
2.2	Architectures et systèmes d'exploitation supportant PVM	39
3.1	Partage de l'espace d'entrée, accélération avec 2 et 4 processeurs . .	58
4.1	Parallélisation modulaire, temps et accélérations en généralisation . .	63
4.2	Temps d'apprentissage et accélérations pour 2 à 10 processeurs . . .	68
4.3	Accélérations comparées	68
4.4	Temps et accélérations de l'algorithme modulaire spécialisé	72
4.5	Pourcentage d'exemples reconnus en fonction du délai (δ)	73
4.6	Nombre de prototypes créés par chaque module pour $\delta = 1$	74
4.7	Nombre de prototypes créés par chaque module pour $\delta > 1$	75
5.1	Latence et taux de transfert sur <i>Ethernet</i>	83
5.2	Latence et taux de transfert sous PVM	92
6.1	Liste des descripteurs proposés au groupe profond GP	103
6.2	Nombres de descripteurs pour les groupes utilisés	103
6.3	Noms des odeurs	104
6.4	Nombres d'exemples, selon les groupes et la nature de la réponse . .	104
6.5	Légende de la table 6.6 des résultats	105
6.6	Résultats de la tentative de classification des odeurs	106
6.7	Classification pour GP avec le nouvel algorithme d'apprentissage . .	111
7.1	Analyse de l'influence de l'amorçage sur l'algorithme de classification	120
7.2	Pourcentage d'exemples reconnus en généralisation	124

A.1 Table de vérité du *ET*, du *OU*, et du *OU exclusif* 139