

Table des matières

Remerciements	i
Résumé	iii
1 Introduction générale	1
1.1 Contexte et problématique	1
1.2 Objectifs et contributions	5
1.3 Organisation de la thèse	8
2 État de l'art	11
2.1 Introduction	11
2.2 Adaptation des données multidimensionnelles	13
2.2.1 Pré-traitement des données multidimensionnelles avec l'OLAP	13
2.2.2 Mise en forme des données multidimensionnelles avec l'OLAP	14
2.2.3 Aplatissage et préparation des données d'un entrepôt	15
2.3 Extension de l'analyse OLAP et des langages de requêtes	16
2.3.1 Fouille de données en ligne	16
2.3.2 Fouille des bases de données multidimensionnelles	18
2.3.3 Utilisation des SGBDMs pour la découverte des connaissances	19
2.3.4 Extension de l'OLAP à la découverte des connaissance	20
2.4 Adaptation des techniques de fouille de données	23
2.4.1 Extraction des connaissances à partir des cubes de données	23
2.4.2 Généralisation des règles d'association aux cubes de données	23
2.4.3 Modèles statistiques dans les cubes de données	25
2.5 Conclusion et positionnement	26
3 Réorganisation des cubes de données par une approche factorielle	33

3.1	Introduction	33
3.2	Objectifs et motivations	36
3.3	Représentation des données multidimensionnelles	38
3.3.1	Compression des cubes de données	38
3.3.2	Organisation des cubes de données	40
3.4	Notations générales	40
3.5	Définitions et formalisation	42
3.5.1	Tableau disjonctif complet	42
3.5.2	Tableau de contingence	43
3.5.3	Équation factorielle	43
3.5.4	Contribution d'une modalité	44
3.5.5	Valeur-test d'une modalité	46
3.6	Arrangement des modalités	47
3.6.1	Arrangement des modalités selon leurs projections	47
3.6.2	Arrangement des modalités selon leurs valeurs-test	48
3.7	Évaluation de la qualité de représentation multidimensionnelle	49
3.7.1	Voisinage d'une cellule dans une représentation multidimensionnelle	49
3.7.2	Similarité entre cellules dans une représentation multidimensionnelle	50
3.7.3	Indice d'homogénéité d'une représentation multidimensionnelle	51
3.7.4	Gain d'homogénéité	53
3.8	Études de cas	53
3.8.1	Cas de l'arrangement des modalités selon leurs projections	54
3.8.2	Cas de l'arrangement des modalités selon leurs valeurs-test	57
3.9	Expérimentations et performances	60
3.10	Conclusion et perspectives	63
4	Agrégation par classification dans les cubes de données	67
4.1	Introduction	67
4.2	Objectifs et motivations	70
4.3	Définitions et formalisation	72
4.3.1	Individus de la classification	72
4.3.2	Variables de la classification	73
4.4	Classification ascendante hiérarchique	75

4.4.1	Mesure de distance entre individus	76
4.4.2	Mesure de distance entre groupes d'individus	77
4.4.3	Algorithme de la classification ascendante hiérarchique	78
4.5	Évaluation des agrégats des modalités	79
4.5.1	Critère des inerties intra et inter-classes	80
4.5.2	Critère de la méthode de <i>Ward</i>	82
4.5.3	Critère de la séparabilité des classes	83
4.6	Conclusion et perspectives	86
5	Explication dans les cubes de données par règles d'association	89
5.1	Introduction	89
5.2	Extraction des règles d'association à partir des cubes de données	91
5.2.1	Historique des règles d'association	91
5.2.2	Règles d'association dans les structures multidimensionnelles	92
5.2.3	Discussion et positionnement	100
5.3	Définitions et formalisation	104
5.3.1	Sous-cube de données	104
5.3.2	Agrégation SUM d'un sous-cube de données	106
5.3.3	Prédicat dimensionnel	106
5.3.4	Prédicat inter-dimensionnels	106
5.4	Méta-règle inter-dimensionnelles	107
5.5	Support et confiance basés sur la mesure	108
5.5.1	Définition classique du support et de la confiance	108
5.5.2	Nouvelle définition du support et de la confiance	109
5.6	Critères de qualité pour les règles inter-dimensionnelles	112
5.6.1	Limites du support et de la confiance	112
5.6.2	Terminologie et notations	112
5.6.3	Choix des indices d'évaluation des règles d'association	113
5.7	Algorithme d'extraction des règles inter-dimensionnelles	115
5.7.1	Terminologie et notations	115
5.7.2	Approches de recherche des motifs fréquents	115
5.7.3	Adaptation d'Apriori aux données multidimensionnelles	117
5.7.4	Calcul des indices des règles inter-dimensionnelles	121
5.8	Visualisation des règles inter-dimensionnelles	122
5.8.1	Principes de visualisation de <i>Bertin</i>	123

5.8.2	Codage graphique d'une règle d'association	123
5.8.3	Visualisation des règles d'association	125
5.9	Expérimentations et performances	127
5.10	Conclusion et perspectives	129
6	Implémentation et cas d'application aux données complexes	133
6.1	Introduction	133
6.2	Plateforme MiningCubes	136
6.2.1	Architecture générale de la plateforme MiningCubes	136
6.2.2	Module de réorganisation par approche factorielle	138
6.2.3	Module d'agrégation par classification	140
6.2.4	Module d'explication par règles d'association	142
6.3	Jeu de données complexes	145
6.3.1	Présentation de la base DDSM	146
6.3.2	Corpus XML des données de mammographies	148
6.4	Méthodologie d'entreposage des données complexes	150
6.4.1	Entrepôts de données XML	150
6.4.2	Vue d'ensemble de la méthodologie X-Warehousing	152
6.5	Construction du cube XML des données de mammographies	153
6.5.1	Contexte d'analyse dans les données de mammographies	154
6.5.2	Cubes XML des données de mammographies	155
6.6	Agrégation des données complexes par classification	158
6.6.1	Objectifs de l'agrégation des données de mammographies par classification	158
6.6.2	Module de connexion aux cubes XML	159
6.6.3	Agrégation par classification dans le cube XML des données de mammographies	159
6.6.4	Évaluation des partitions des agrégats	163
6.7	Expérimentations et performances	165
6.8	Conclusion et perspectives	167
7	Vers un cadre formel général	171
7.1	Introduction	171
7.2	Modèles multidimensionnels et algèbres OLAP	172
7.2.1	Algèbre pour les matrices bi-dimensionnelles	173
7.2.2	Algèbre pour les cubes de données	174

7.2.3	Algèbres pour les cubes de données relationnels	175
7.2.4	Algèbre pour les informations réparties en niveaux	175
7.2.5	Algèbre pour les cubes vus comme des ensembles de relations .	176
7.2.6	Algèbre pour l'analyse en ligne OLAP	177
7.2.7	Algèbre pour les tables multidimensionnelles	177
7.2.8	Discussion et positionnement	178
7.3	Espace de données multidimensionnelles	182
7.3.1	Espace de caractéristiques	183
7.3.2	Espace d'attributs	183
7.3.3	Espace de modalités	185
7.3.4	Ensemble de cellules	187
7.3.5	Ensemble de fonctions d'agrégation	188
7.4	Noyau minimal fermé d'une algèbre multidimensionnelle	189
7.4.1	Opérateurs de structuration	189
7.4.2	Opérateurs de navigation	192
7.5	Vers une extension à la fouille de données en ligne	195
7.5.1	Réorganisation par une ACM (ORCA)	196
7.5.2	Agrégation par une CAH (OPAC)	198
7.6	Conclusion et perspectives	200
8	Conclusion générale	203
8.1	Bilan et contributions	203
8.2	Perspectives de recherche	205
	Bibliographie	207
	Liste des figures	224
	Liste des tableaux	228
	Liste des algorithmes	231
	Table des matières	232