

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université HADJ LAKHDAR – BATNA
Faculté des sciences de l'ingénieur
Département d'informatique

N° d'ordre :
Série :

Mémoire

En vue de l'obtention du diplôme de Magister en informatique

Spécialité : *Sciences et Technologies de l'Information et de la Communication (STIC)*

Option : *Systèmes d'Information et de Connaissance (SIC)*

Une plate forme orientée agent pour le data mining

Par :

M^{lle}. CHAMI Djazia

Présenté le : / /

Devant le jury composé de :

Président : Dr. BELATTAR Brahim, Maître de Conférence à l'université de Batna.

Rapporteur : Dr. KAZAR Okba, Maître de Conférence à l'université de Biskra.

Examineur : Dr. ZIDANI Abdelmadjid, Maître de Conférence à l'université de Batna.

Examineur : Dr. BILAMI Azeddine, Maître de Conférence à l'université de Batna.

2009-2010

Résumé :

Le Data Mining est une technologie dont le but est la valorisation de l'information et l'extraction de connaissances d'un grand nombre de données. Cette technologie qui est devenue un outil important pour améliorer les revenus des entreprises rencontre des problèmes dus à la quantité énorme de données à exploiter. Pour y parvenir, nous avons fait appel au paradigme de système multi agents pour distribuer la complexité sur plusieurs entités autonomes appelés agents. Le résultat de notre recherche est une modélisation d'un système de clustering basé agents. Notre approche qui est une approche cognitive se base sur la connaissance pour assurer le bon fonctionnement du système et sa fiabilité.

Mots clés: data mining, système multi-agents.

ملخص:

تنقيب البيانات هي تكنولوجيا الهدف منها تجميع المعلومات و استخراج المعرفة من عدد كبير من البيانات. هذه التكنولوجيا التي أصبحت أداة هامة لتحسين أرباح الشركات تعاني من مشاكل نظرا للكمية الهائلة من البيانات الواجب تنقيبها. من اجل معالجة هذه المشاكل استعنا بالأنظمة متعددة الوكلاء من أجل توزيع المهام على عدد من الكيانات المستقلة. نتيجة بحثنا هي بناء نظام متعدد الوكلاء لتجميع البيانات. منهجنا منهج إدراكي قائم على المعرفة لضمان الأداء السليم للنظام وموثوقيته .

الكلمات الرئيسية: تنقيب البيانات، الأنظمة متعددة الوكلاء.

SOMMAIRE

INTRODUCTION GENERALE	1
------------------------------------	----------

CHAPITRE I: GENERALITE SUR LE DATA MINING

1 INTRODUCTION	4
-----------------------------	----------

2 DEFINITION DU DATA MINING	4
--	----------

3 MOTIVATIONS.....	5
---------------------------	----------

4 DATA MINING SUR QUELS TYPES DE DONNEES ?	8
---	----------

4.1 LES FICHIERS PLATS	8
------------------------------	---

4.2 LES BASES DE DONNEES RELATIONNELLES	8
---	---

4.3 LES DATA WAREHOUSES.....	8
------------------------------	---

4.4 LES BASES DE DONNEES TRANSACTIONNELLES.....	10
---	----

4.5 LES BASES DE DONNEES MULTIMEDIA	11
---	----

4.6 LES BASES DE DONNEES SPATIALES.....	11
---	----

4.7 LES BASES DE DONNEES DE SERIES TEMPORELLES	11
--	----

4.8 LE WORLD WIDE WEB	12
-----------------------------	----

5 LES TACHES DU DATA MINING	13
--	-----------

5.1 LA CLASSIFICATION	13
-----------------------------	----

5.2 L'ESTIMATION	14
------------------------	----

5.3 LA PREDICTION	14
-------------------------	----

5.4 LE GROUPEMENT PAR SIMILITUDE	15
--	----

5.5 L'ANALYSE DES CLUSTERS	15
----------------------------------	----

5.6 LA DESCRIPTION	15
--------------------------	----

6 LES ETAPES DU PROCESSUS DE DATA MINING	16
---	-----------

7 TECHNIQUES DU DATA MINING	17
--	-----------

7.1 LES RESEAUX DE NEURONES	17
-----------------------------------	----

7.1.1 AVANTAGES ET INCONVENIENTS	18
--	----

7.2 LES ARBRES DE DECISION	18
----------------------------------	----

7.2.1 LES ALGORITHMES D'INDUCTION DES ARBRES DE DECISION	19
--	----

7.2.2 AVANTAGES ET INCONVENIENTS	19
--	----

7.3 LES ALGORITHMES GENETIQUES	19
--------------------------------------	----

7.3.1 PRINCIPE DE BASE DES ALGORITHMES GENETIQUES	20
---	----

7.3.2 CODAGE D'UN ALGORITHME GENETIQUE	20
--	----

7.3.3	AVANTAGES ET INCONVENIENTS	21
7.4	LES REGLES ASSOCIATIVES	22
7.4.1	LES ALGORITHMES D'INDUCTION DES REGLES ASSOCIATIVES.....	22
7.4.2	AVANTAGES ET INCONVENIENTS	22
7.5	L'ALGORITHME DES K-PLUS PROCHES VOISINS	23
7.5.1	ALGORITHME DE CLASSIFICATION PAR K-PPV.....	23
7.5.2	COMMENT CELA MARCHE-T-IL ?	23
7.5.3	AVANTAGES ET INCONVENIENTS	24
7.6	L'ALGORITHME DES K-MOYENNES (K-MEANS).....	25
7.6.1	PRINCIPE DE FONCTIONNEMENT	25
7.7	ALGORITHME DE CLUSTERING PAR K-MEANS.....	25
7.7.1	EVALUATION	25
7.7.2	AVANTAGES ET INCONVENIENTS	26
8	<u>CATEGORISATION DES SYSTEMES DU DATA MINING.....</u>	<u>26</u>
9	<u>DOMAINES D'APPLICATION DU DATA MINING</u>	<u>27</u>
9.1	LE DATA MINING DANS LE SECTEUR BANCAIRE	27
9.2	LE DATA MINING DANS LA BIO-INFORMATIQUE ET LA BIOTECHNOLOGIE.....	28
9.3	LE DATA MINING DANS LE MARKETING DIRECT ET LE COLLECTE DE FONDS.....	28
9.4	LE DATA MINING DANS LA DETECTION DE FRAUDE	29
9.5	LE DATA MINING DANS LA GESTION DE DONNEES SCIENTIFIQUES	29
9.6	LE DATA MINING DANS LE SECTEUR DES ASSURANCES	30
9.7	LE DATA MINING DANS LA TELECOMMUNICATION.....	30
9.8	LE DATA MINING DANS LA MEDECINE ET LA PHARMACIE.....	30
9.9	LE DATA MINING DANS LE COMMERCE AU DETAIL	31
9.10	LE DATA MINING DANS LE E-COMMERCE ET LE WORLD WIDE WEB	31
9.11	LE DATA MINING DANS LE MARCHE BOURSIER ET L'INVESTISSEMENT.....	31
9.12	LE DATA MINING DANS L'ANALYSE DE CHAINE D'APPROVISIONNEMENT.....	32
10	<u>CONCLUSION.....</u>	<u>32</u>
<u>CHAPITRE II: ÉTUDE DES TRAVAUX EXPLOITANT LES SMAS POUR LE DATA MINING</u>		
1	<u>INTRODUCTION</u>	<u>34</u>
2	<u>L'INTELLIGENCE ARTIFICIELLE DISTRIBUEE (IAD).....</u>	<u>34</u>
3	<u>CONCEPT D'AGENT</u>	<u>35</u>
3.1	DEFINITIONS.....	35
3.2	DIFFERENCE ENTRE OBJET ET AGENT.....	36
3.3	TYPES D'AGENTS	36
3.3.1	LES AGENTS COGNITIFS.....	36
3.3.2	LES AGENTS REACTIFS.....	37
4	<u>LES SYSTEMES MULTI-AGENTS.....</u>	<u>37</u>

4.1	DEFINITIONS	37
4.2	QUAND UTILISER UN SMA?	38
5	<u>QUELQUES TRAVAUX EXPLOITANT LES SMAS POUR LE DM</u>	<u>38</u>
5.1	APPROCHE SMA POUR LA SEGMENTATION MARKOVIENNE DES TISSUS ET STRUCTURES PRESENTS DANS LES IRM CEREBRALES.....	38
5.1.1	IMPLEMENTATION SMA.....	39
5.1.2	EVALUATION SUR IMAGES REELLES ACQUISES A 3T.....	43
5.2	UNE APPROCHE SMA DE L'AGREGATION ET DE LA COOPERATION DES CLASSIFIEURS	43
5.2.1	UNE APPROCHE FONDEE SUR LES SMA	44
5.2.2	FONCTIONNEMENT.....	44
5.2.3	SYNTHESE	46
5.3	UNE APPROCHE POUR L'EXTRACTION DES REGLES D'ASSOCIATION SPATIALES BASEE MULTI-AGENT : RASMA	47
5.3.1	ARCHITECTURE DE RASMA	50
5.3.2	PRESENTATION DES AGENTS.....	51
5.3.3	EXPERIMENTATION, RESULTATS ET PERFORMANCES	55
5.3.4	EVALUATION	56
6	<u>CONCLUSION.....</u>	<u>57</u>
<u>CHAPITRE III: MODÉLISATION</u>		
1	<u>INTRODUCTION.....</u>	<u>59</u>
2	<u>PRESENTATION GENERALE DU MODELE.....</u>	<u>59</u>
3	<u>DESCRIPTION DES COMPOSANTS DE L'ARCHITECTURE</u>	<u>61</u>
3.1	AGENT INTERFACE	61
3.1.1	ARCHITECTURE DE L'AGENT INTERFACE	61
3.1.2	FONCTIONNEMENT DE L'AGENT INTERFACE.....	62
3.1.3	LE SAVOIR DE L'AGENT INTERFACE.....	62
3.2	AGENT INIT-CLUSTER	63
3.2.1	ARCHITECTURE DE L'AGENT INIT-CLUSTER	63
3.2.2	FONCTIONNEMENT DE L'AGENT INIT-CLUSTER	64
3.2.3	LE SAVOIR DE L'AGENT INIT-CLUSTER.....	64
3.3	AGENT AFFECT-CLUSTER	64
3.3.1	ARCHITECTURE DE L'AGENT AFFECT-CLUSTER	65
3.3.2	FONCTIONNEMENT DE L'AGENT AFFECT-CLUSTER.....	66
3.3.3	LE SAVOIR DE L'AGENT AFFECT-CLUSTER	67
3.4	AGENT CALC-CENTROÏDE	68
3.4.1	ARCHITECTURE DE L'AGENT CALC-CENTROÏDE.....	68
3.4.2	FONCTIONNEMENT DE L'AGENT CALC-CENTROÏDE	69
3.4.3	LE SAVOIR DE L'AGENT CALC-CENTROÏDE	69
3.5	AGENT CLAC-DISTANCE	70
3.5.1	ARCHITECTURE DE L'AGENT CALC-DISTANCE	70
3.5.2	FONCTIONNEMENT DE L'AGENT CALC-DISTANCE	71
3.5.3	LE SAVOIR DE L'AGENT CALC-DISTANCE.....	71

4	<u>LA COMMUNICATION INTER-AGENTS.....</u>	<u>72</u>
5	<u>FONCTIONNEMENT DU SYSTEME</u>	<u>73</u>
5.1	DIAGRAMME DE CLASSES	73
5.2	DIAGRAMME DE SEQUENCE	73
6	<u>CONCLUSION.....</u>	<u>75</u>
<u>CHAPITRE IV: ÉTUDE DE CAS</u>		
1	<u>INTRODUCTION</u>	<u>77</u>
2	<u>LES ETAPES DU PROCESSUS ADOPTE</u>	<u>77</u>
2.1	PREPARATION DES DONNEES.....	77
2.2	NETTOYAGE ET TRANSFORMATION DES DONNEES	79
2.3	DATA MINING	84
3	<u>OUTILS DE PROGRAMMATION</u>	<u>90</u>
3.1	POURQUOI JAVA?.....	90
3.2	LA PLATEFORME JADE	90
3.2.1	L'ENVIRONNEMENT D'EXECUTION JADE	92
3.2.2	LA COMMUNICATION ENTRE LES AGENTS JADE	92
4	<u>CONCLUSION.....</u>	<u>93</u>
	<u>CONCLUSION GENERALE</u>	<u>95</u>
	<u>REFERENCES</u>	<u>97</u>