

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Ecole nationale Supérieure d'Informatique (E.S.I)

Mémoire

**En vue de l'obtention du diplôme de Magister
En Informatique**

Option

**Ingénierie des Systèmes Informatiques
(ISI)**

Présenté par

Houda OUFAIDA

Thème :

**Amélioration du processus de démarrage à froid
dans les systèmes de filtrage d'information
collaboratif**

Soutenu devant le jury :

M. Amar	BALLA	Président	Maître de conférences (ESI)
M. Khaled-Walid	HIDOUCI	Examineur	Maître de conférences (ESI)
M. Reda .A	GHOMARI	Examineur	Maître de conférences (ESI)
M. Omar	NOUALI	Directeur de mémoire	Maitre de Recherche (CERIST)

2008-2009

Résumé

Le filtrage d'information est le processus permettant à partir d'un large volume d'informations dynamiques, d'extraire les informations intéressants un utilisateur ou un groupe d'utilisateurs. Cependant, ce type de systèmes souffre toujours d'un certain nombre de limitations, notamment, la rareté des votes et le démarrage à froid pour un nouvel utilisateur, un nouveau document ou un nouveau système.

La modélisation du profil utilisateur est un élément clé dans tout système de filtrage d'information. La collecte et l'organisation des données relatives à un utilisateur s'avèrent critiques pour le processus de recommandation surtout dans des situations de démarrage à froid.

Dans le but d'atténuer les effets du démarrage à froid, nous avons proposé une modélisation multidimensionnelle du profil utilisateur constituée de trois dimensions : collaborative, sociale et sémantique. Chacune de ces dimensions est ensuite exploitée par un moteur de génération de recommandations. Ces recommandations sont ainsi classées et présentées à l'utilisateur. Une expérimentation est alors conduite pour valider notre proposition.

Mots clés : Filtrage d'information, Filtrage d'information collaboratif, démarrage à froid, profil utilisateur, Web sémantique.

Abstract

Recommendation process aims to provide users with interesting information selected from large and dynamic information sources. However, these systems still have some limitations such as sparsity problem and cold start problem for a new user, a new document and both.

User modeling is a key element for the success of every recommendation system. User's information collection and organization are critical, especially in cold start situations.

Thus, we propose a multi view recommendation engine integrating, in addition of the collaborative recommendations, social and semantic recommendations. Three different hybridization strategies are also proposed to combine different types of recommendations. An empirical study was conducted to verify our proposition.

Keywords

Recommender systems, collaborative filtering, cold start problem, user modeling, semantic web.

Sommaire

Introduction générale	1
Partie 1 Etat de l'art	3
Chapitre 1 Filtrage d'Information.....	4
1.Filtrage d'information	5
1.1.Systèmes de filtrage à base de contenu	6
1.2.Systèmes de filtrage collaboratif	8
1.2.1.Filtrage basé mémoire	9
1.2.2.Filtrage basé modèle.....	11
1.3.Systèmes de filtrage hybride	13
2.Filtrage collaboratif et le web 2.0.....	14
2.1.Systèmes de filtrage d'information à base d'ontologies	14
2.2.Systèmes d'annotations collaboratives.....	16
2.3.Systèmes de filtrage d'informations à base de réseaux sociaux.....	17
3.Exemples de systèmes de filtrage collaboratif	19
3.1.Tapestry	19
3.2.GroupLens	21
3.3.Le système de Malltz et Ehrlich	22
3.4.CoCoFil	22
3.5.Amazon	23
4.Evaluation des systèmes de filtrage d'information	23
5.Synthèse des approches de filtrage d'information	25
6.Conclusion.....	27
Chapitre 2 Démarrage à Froid.....	28
1.Problématique du Démarrage à Froid.....	29
1.1.Démarrage à froid pour un nouveau document	30
1.2.Démarrage à froid pour un nouvel utilisateur	30
1.3.Démarrage à froid pour un nouveau système	30
2.Solutions existantes	31
2.1.Filtrage collaboratif actif	31
2.2.Approche conversationnelle	31
2.3.Recommandations exploratoires	32

2.4.Approche sociodémographique.....	33
2.5.Approches probabilistes	34
2.6.Approche à base d'agents.....	37
2.7.Approches par inférence.....	38
2.8.Approches Ontologiques	40
2.9.Approches par hybridation	43
3.Conclusion.....	44
Chapitre 3 Profil Utilisateur	45
1.Définition du profil utilisateur.....	46
2.Caractéristiques du profil utilisateur	47
2.1.Exemples de documents	47
2.2.Caractéristiques de documents	47
2.3.Attributs sur l'utilisateur	47
2.4.Règles et formules de préférences.....	48
3.Acquisition du Profil Utilisateur	48
3.1.Acquisition explicite	49
3.2.Acquisition implicite.....	49
4.Evolution du Profil Utilisateur	50
4.1.Approche statistique	50
4.2.Approche collaborative	51
4.3.Approche par apprentissage	51
5.Représentation du profil utilisateur	53
5.1.Représentation vectorielle	53
5.2.Représentation sémantique.....	54
5.3.Représentation multidimensionnelle	54
6.Exemples de profils	58
7.Conclusion.....	59
Partie 2 Proposition et Validation	61
Chapitre 4 Approche proposée	62
1.Proposition.....	64
1.1.Architecture globale	64
1.2.Modélisation du profil utilisateur.....	65
1.2.1.Dimension collaborative	65

1.2.2. Dimension sociale	66
1.2.3. Dimension sémantique	67
1.3. Moteurs de recommandation	69
1.3.1. Vote Matching	70
1.3.2. Social Matching	70
1.3.3. Concept Matching	72
1.4. Stratégies d'hybridation	79
1.4.1. Hybridation Simple	81
1.4.2. Hybridation par combinaison	82
1.4.3. Hybridation par commutation	82
2. Conclusion	83
Chapitre 5 Prototype et Évaluation.....	85
1. Le jeu de données « BookCrossing »	86
1.1. Données évaluations	86
1.2. Données démographiques	87
1.3. Données Livres	88
2. Données sémantiques	88
2.1. Amazon Associates Web Service	88
2.2. Indexation des livres	90
3. Génération des profils utilisateur	91
3.1. Construction de la dimension collaborative	91
3.2. Construction de la dimension sociale	92
3.3. Construction de la dimension sémantique	92
3.3.1. Similarité sémantique entre catégories	93
4. Evaluation	95
4.1. Métriques d'évaluation	95
4.2. Scénarios d'évaluations	96
4.3. Résultats	96
5. Conclusion	105
Conclusion et perspectives	106
Annexe A Web Sémantique.....	110
A.1. Définition	110
A.2. Technologies du Web Sémantique	111

A.3. Ontologies.....	112
A.4. Exemples d'ontologies.....	114
A.5. Similarité sémantique	114
Annexe B Amazon Web Services.....	121
B.1. Amazon Associates Web Service (A2S)	121
B.2. Simple Storage Service (S3).....	124
B.3. Elastic Compute Cloud (EC2)	124
Bibliographie	125