

École Nationale Supérieure d'Informatique et d'Analyse des Systèmes
Centre d'Études Doctorales en Sciences des Technologies de l'Information et de l'Ingénieur

AVIS DE SOUTENANCE DE THÈSE DE DOCTORAT

Monsieur Bihi SABIRI

Soutiendra publiquement sa thèse de Doctorat en Informatique

Le 07 octobre 2026 à 9h30 au Grand Amphi à l'ENSIAS

Intitulé de la thèse

HYBRID QUALITY BASED RECOMMENDATION SYSTEMS FOR DISSEMINATION OF INFORMATION IN BIG DATA

Président :

Pr. Ahmed ETTALBI, PES, ENSIAS, Université Mohammed V de Rabat

Directeur de thèse :

Pr. Mahmoud NASSAR, PES, ENSIAS, Université Mohammed V de Rabat

Co-encadrant de thèse (invité) :

Pr. Amal KHTIRA, MC, ESTS, Université Mohammed V de Rabat

Rapporteurs :

Pr. Samir MBARKI, PES, Faculté des Sciences, Université Ibn Tofail de Kénitra

Pr. Abdessamad BELANGOUR, PES, FSBM, Université Hassan II de Casablanca

Pr. Yasser EL MADANI EL ALAMI, MCH, ENSIAS, Université Mohammed V de Rabat

Examineur :

Pr. Widad ETTAZI, MCH, ENSIAS, Université Mohammed V de Rabat

Abstract: In the era of big data, the rapid increase in information volumes, driven by the proliferation of mobile devices and constant internet access, has fundamentally altered the manner in which users access information. Substantial quantities of data are generated daily; however, their quality and relevance vary according to users' needs and priorities, thereby underscoring the necessity for effective recommendation systems. Conventional approaches, such as collaborative filtering and content-based filtering, have yielded notable advancements, yet still exhibit significant limitations, including data sparsity, overfitting, the absence of strategic modelling, and insufficient semantic understanding of heterogeneous inputs.

In this context, this thesis proposes HyQARF (Hybrid Quality-Aware Recommender Framework), a theoretically grounded architecture designed to address these limitations through the integration of complementary technologies and algorithmic paradigms. Semantic representation is achieved through a set of techniques spanning Large Language Models, Natural Language Processing methods, and contrastive learning, selected and combined according to the nature of the input data and the modality of the application domain. The recommendation ecosystem is then conceptualized as a strategic interaction space, where Nash Equilibrium ensures that no agent can unilaterally improve its outcome, thereby guaranteeing the stability and consistency of the produced recommendations. Fair and interpretable assessment of each agent's contribution to the collective outcome is achieved through the Shapley value, which accounts for all possible collaboration configurations to produce objective quality scores. Finally, the combinatorial complexity inherent in large-scale recommendation scenarios is addressed through Ant Colony Optimisation, which guides the search within complex solution spaces by exploiting accumulated experience to navigate non-convex optimisation landscapes efficiently.

The framework is validated across two structurally distinct application domains. In the domain of professional profile matching, HyQARF is applied to the automatic matching of candidate profiles and mission offers in a real-world recruitment context, where the multi-algorithmic pipeline is deployed on a scalable distributed architecture designed to handle large volumes of candidate profiles and job offers efficiently. This study is based on a distributed architecture that leverages the Hadoop ecosystem for the storage and processing of massive datasets, in conjunction with the Mahout library for the implementation of large-scale machine learning algorithms.

In the domain of image generation, Generative Adversarial Networks are leveraged as a means of addressing data scarcity in recommendation pipelines by generating synthetic content, with the adversarial training dynamics formally analysed through the game-theoretic and optimisation mechanisms of HyQARF. Both implementations are rigorously evaluated, providing component-level diagnostics that identify current limitations and formulate targeted directions for future improvement.

The experimental results demonstrate the effectiveness of HyQARF in addressing the principal limitations of existing hybrid recommendation systems, while providing a principled, modular, and extensible architecture applicable across heterogeneous domains and data modalities.

Keywords: Strategy recommendation systems, Hybrid quality-based recommendations, Big data, Machine learning, Neural network.

Résumé : À l'ère des données massives, l'augmentation rapide des volumes d'information, portée par la prolifération des appareils mobiles et l'accès permanent à internet, a fondamentalement modifié la manière dont les utilisateurs accèdent à l'information. Des quantités substantielles de données sont générées quotidiennement ; cependant, leur qualité et leur pertinence varient selon les besoins et les priorités des utilisateurs, soulignant ainsi la nécessité de systèmes de recommandation efficaces. Les approches conventionnelles, telles que le filtrage collaboratif et le filtrage basé sur le contenu, ont permis des avancées notables, mais présentent encore des limitations significatives, notamment la parcimonie des données, le surapprentissage, l'absence de modélisation stratégique et une compréhension sémantique insuffisante des entrées hétérogènes.

Dans ce contexte, cette thèse propose HyQARF (Hybrid Quality-Aware Recommender Framework), une architecture théoriquement fondée et conçue pour répondre à ces limitations par l'intégration de technologies et de paradigmes algorithmiques complémentaires. La représentation sémantique est assurée par un ensemble de techniques couvrant les grands modèles de langage, les méthodes de traitement automatique du langage naturel et l'apprentissage contrastif, sélectionnés et combinés selon la nature des données d'entrée et la modalité du domaine d'application. L'écosystème de recommandation

est ensuite conceptualisé comme un espace d'interaction stratégique, où l'équilibre de Nash garantit qu'aucun agent ne peut améliorer unilatéralement son résultat, assurant ainsi la stabilité et la cohérence des recommandations produites. Une évaluation juste et interprétable de la contribution de chaque agent au résultat collectif est réalisée par la valeur de Shapley, qui tient compte de toutes les configurations de collaboration possibles pour produire des scores de qualité objectifs. Enfin, la complexité combinatoire inhérente aux scénarios de recommandation à grande échelle est traitée par l'optimisation par colonies de fourmis, qui guide la recherche dans des espaces de solutions complexes en exploitant l'expérience accumulée pour naviguer efficacement dans des paysages d'optimisation non convexes.

Le cadre est validé sur deux domaines d'application structurellement distincts. Dans le domaine de la mise en correspondance de profils professionnels, HyQARF est appliqué à l'appariement automatique de profils de candidats et d'offres de missions dans un contexte de recrutement réel, où le pipeline multi-algorithmique est déployé sur une architecture distribuée et évolutive conçue pour traiter efficacement de grands volumes de profils et d'offres d'emploi. En effet, cette étude exploite une architecture distribuée basée sur l'écosystème Hadoop pour le traitement et la gestion de volumes considérables de données, combinée à la bibliothèque Mahout pour l'exécution d'algorithmes d'apprentissage automatique à grande échelle. Dans le domaine de la génération d'images, les réseaux antagonistes génératifs sont exploités comme moyen de pallier la parcimonie des données dans les pipelines de recommandation en générant du contenu synthétique, tandis que la dynamique d'entraînement adversarial est formellement analysée à travers les mécanismes de la théorie des jeux et d'optimisation de HyQARF. Les deux implémentations font l'objet d'une évaluation rigoureuse, fournissant des diagnostics au niveau des composants qui identifient les limitations actuelles et formulent des directions ciblées pour les travaux futurs.

Les résultats expérimentaux démontrent l'efficacité de HyQARF pour répondre aux principales limitations des systèmes de recommandation hybrides existants, tout en proposant une architecture modulaire, fondée et extensible, applicable à des domaines et des modalités de données hétérogènes.

Mots-clés : Systèmes de recommandation stratégique, Recommandations hybrides basées sur la qualité, Mégadonnées, Apprentissage automatique, Réseau neuronal.