

(12) BREVET D'INVENTION

- (11) N° de publication : **MA 50776 B1**
- (51) Cl. internationale : **G06F 17/00; G06F 17/30; G06F 17/00; G06F 17/30**
- (43) Date de publication : **30.06.2022**
-
- (21) N° Dépôt : **50776**
- (22) Date de Dépôt : **03.09.2020**
- (71) Demandeur(s) : **Université Mohammed V de Rabat, Angle avenue Allal El Fassi et Mfadel Cherkaoui Al Irfane , Rabat, 8007 (MA)**
- (72) Inventeur(s) : **EL HANDRI Kaoutar ; IDRISSE Abdellah**
- (74) Mandataire : **Kartit Zaid**
-
- (54) Titre : **System collaboratif d'aide à la décision à base des recommandations multi critères**
- (57) Abrégé : L'invention présentée dans ce document montre un système de recommandation combinant le domaine MCDA, et les techniques de traitement des requêtes dominantes de TOP-k, en particulier les outils MCDA utilisés pour modéliser les préférences des utilisateurs, ainsi que la technique de filtrage collaboratif pour identifier les éléments inconnus préférés par chaque utilisateur de système. Nous démontrons cette méthodologie proposée comme un système de recommandation distribué nouveau qui se base sur des techniques adaptés et améliorés. De plus, par une étude comparative avec d'autres approches utilisant différents algorithmes TOPk , l'application SPTOP-kWS sur des bases de données réelles et synthétiques prouve que la parallélisation de cet algorithme utilisant efficacement le modèle de programmation de haut niveau fourni par Apache Spark et basé sur le SVD de Funk fait partie intégrante d'un processus de recommandation amélioré.

Abrégé

L'invention présentée dans ce document montre un système de recommandation combinant le domaine MCDA, et les techniques de traitement des requêtes dominantes de TOP-k, en particulier les outils MCDA utilisés pour modéliser les préférences des utilisateurs, ainsi que la technique de filtrage collaboratif pour identifier les éléments inconnus préférés par chaque utilisateur de système. Nous démontrons cette méthodologie proposée comme un système de recommandation distribué nouveau qui se base sur des techniques adaptés et améliorés. De plus, par une étude comparative avec d'autres approches utilisant différents algorithmes TOP-k, l'application SPTOP-kWS sur des bases de données réelles et synthétiques prouve que la parallélisation de cet algorithme utilisant efficacement le modèle de programmation de haut niveau fourni par Apache Spark et basé sur le SVD de Funk fait partie intégrante d'un processus de recommandation amélioré.

Titre : System collaboratif d'aide à la décision à base des recommandations multi critères**DOMAINE DE L'INVENTION**

La présente invention concerne un système d'Aide à la Décision Multicritère (MCDA pour Multi Criteria Decision Aiding) appliquée dans le domaine Big Data. Il s'agit plus précisément d'une méthode de traitement et d'optimisation de requête multicritères pour générer des réponses multi-objectifs qui facilitent la prise de bonne décision.

Il existe de nombreuses méthodes mathématiques d'analyse multicritères pour générer un ensemble d'éléments multi-objectifs. Ces méthodes peuvent être regroupées en deux approches. Une première approche basée sur la théorie de l'utilité qui utilise des critères d'agrégation à priori en un seul test, et une seconde approche basée sur des méthodes de classement telles que ELECTRE I, III, IV, IS, PROMETHEE I et PROMETHEE II. L'usage de ces méthodes est fortement relié au domaine de filtrage collaboratif qui est l'approche la plus générale pour concevoir un système de recommandation.

Il est connu dans la littérature que les algorithmes Top-k constituent une exigence cruciale dans de nombreux environnements interactifs, tels que la recherche des services Cloud, les méta-moteurs de recherche sur le Web, les réseaux sociaux, le multimédia et dans tous les types de systèmes de recommandation.

Dans la présente invention, le traitement de requêtes Top-k, consiste à trouver les k-objets qui ont les meilleurs scores globaux. Une requête dans une base de données combine différents attributs notés par une fonction score appelé aussi une fonction d'agrégation. La notation globale de chaque objet sera calculée par cette fonction, et nous pouvons retourner les meilleurs objets Top-k.

Dans le domaine de Big data le volume de données est important ainsi que la vitesse d'arriver de ces données. Par conséquent les systèmes de recommandation dans ce domaine ne supportent pas les opérations de classement dans un environnement distribué. Cela exige des solutions parallèles efficaces en termes de coût et de qualité pour s'adapter aux ressources restreintes.

En outre, différentes techniques de Big Data sont utilisées pour résoudre ce problème. De même, la transition vers les processeurs multi-cœurs, qui est due aux limitations des

processeurs mono cœurs, nécessite des calculs parallèles pour améliorer les performances des logiciels ce qui est l'objectif principale de la parallélisations de l'algorithme au sein d'un système de recommandation distribué proposé dans la présente invention.

ART ANTERIEUR

Plusieurs articles ont abordé la question du traitement des requêtes Top k dans les systèmes de gestion de bases de données centralisées. Dans le papier (Fagin, Ronald, Lotem, A., Naor, M., 2003. Optimal agrégation algorithms for middleware), l'auteur a donné des algorithmes de Top-k intéressants comme les algorithmes : Fagin Algorithm (FA), Threshold Algorithm (TA) et No Random access Algorithm (NRA). Ensuite, il a fourni un algorithme plus amélioré présenté dans Ilyas, I., Beskales, G., A. Soliman, M., 2008. A Survey of Top-k Query Processing Techniques in Relational Database Systems. /doi.org/10.1145/1391729.1391730.

En effet, le FA et le TA sont destinés à des scénarios où les coûts d'accès triés et aléatoires sont comparables. En revanche, l'algorithme NRA est utilisé pour traiter les cas où l'accès aléatoire est coûteux ou impossible.

D'autre part, les approches les plus intéressantes à la question du traitement des requêtes Top-k ont été proposées par Ilyas, I., Beskales, G., A. Soliman, M., 2008. A Survey of Top-k Query Processing Techniques in Relational Database Systems. <https://doi.org/10.1145/1391729.1391730> ; dans ces études, l'auteur a présenté l'état de l'art pour distinguer les différentes catégories des algorithmes Top-k. Et il montre que dans les mêmes cas, les algorithmes Top-k peuvent être efficaces et ingénieux si l'espace de recherche devient immense.

Dans une étude récente basée sur une approche décentralisée présentés dans Amagata, D., Hara, T. and Onizuka, M. (2018) 'Space filling approach for distributed processing of top-k dominating queries'. Les auteurs ont conçu l'algorithme de remplissage de l'espace (SFA : Space Filling Algorithm) et ont utilisé l'approche des points virtuels, pour le traitement de requête de dominance.

Cet algorithme est basé sur une fonction parallèle de calcul des préférences pondérées. Cependant, malgré le fait que les auteurs ont présenté des améliorations pour cet algorithme, à

savoir l'algorithme de la SFA approximative (A-SFA : Approximate SFA), et l'algorithme de la SFA basée sur l'échantillonnage (S-SFA : Sampling-based SFA) pour réduire davantage le temps de traitement des requêtes, les auteurs n'utilisent pas de preuve de profondeur de l'efficacité dimensionnelle de l'algorithme, particulièrement dans plus de cinq dimensions. Malgré cela, les améliorations peuvent fournir un ensemble approximatif de réponses et améliore l'efficacité du traitement des requêtes même si cet objectif a été atteint en sacrifiant la qualité de précision.

De plus, contrairement à notre approche qui utilise une fonction de score de somme pondérée adaptée pour l'optimisation multi-objectif, l'approche discutée dans Amagata et al. (2018) cité auparavant, utilise une méthode traditionnelle de somme pondérée en anglais Weighted Sum Method (WSM) pour traiter le problème de Skyline (par définition c'est un opérateur qui renvoie un ensemble de points qui ne sont pas dominés par d'autres points dans les jeux de données). Selon la définition des requêtes Skyline un service par exemple qui n'est dominé par aucun autre service est dit être un service Skyline ou se trouver dans le Skyline. Les services qui se trouvent dans le Skyline sont les meilleurs compromis possibles entre tous les critères qui intéressent l'utilisateur final dans un système de recommandation. Les Services peuvent être un choix d'un traitement parmi plusieurs pour un médecin ou un choix d'un Cloud Computing pour une entreprise, etc.

La plupart des études visant à affiner le résultat de Skyline constatent que l'interrogation d'une donnée multidimensionnelle peut aboutir soit à un très grand ou à un insuffisant nombre de réponses. Dans les deux cas, cela peut brouiller le choix de l'utilisateur final.

Pour faire face à ces problèmes, différentes approches ont été introduites pour affiner et réduire la taille de Skyline en sélectionnant les points les plus intéressants. En tant que concept commun dans l'optimisation multi-objective, la minimisation d'une somme pondérée est une méthode indépendante ou un composant des autres méthodes. Par conséquent, la compréhension des caractéristiques de cette méthode a de profondes implications. Cependant, malgré les nombreuses applications publiées, il y a une discussion un peu plus détaillée sur la signification conceptuelle des poids et des techniques qui peuvent améliorer l'efficacité de WSM.

Ces travaux précédents n'ont pas utilisé un TOP-k basé sur la méthode de la somme pondérée adaptée pour affiner le résultat du Skyline. Cette approche de base peut être considérée comme très puissante, car dans notre cas nous combinons les avantages de la requête TOP-k et de la méthode (MCDA), à la requête Skyline. De plus, la WSM adaptée nous a permis de modéliser le problème dans un contexte multi-objectif. Le choix de TOP-k comme étape de post-traitement du résultat du Skyline a été motivé par la relation existante entre le Skyline et la méthode TOP-k. On a observé que l'objet top-1 (le traitement choisi par le médecin de l'exemple précédent) pour toute fonction de plus en plus monotone appartient à l'ensemble de Skyline tel que décrit et prouvé dans le travail : Saouk, M., Doulkeridis, C., Vlachou, A., Nørsvag, K., 2016. Efficient processing of TOP-k Joins in MapReduce. Proc IEEE Big Data.

En effet, presque tous les principaux algorithmes de TOP-k comme ceux de Fagin souffrent de comportements statiques, et donc ne s'adaptent pas aux coûts d'accès à l'exécution. Dans certains cas, leurs heuristiques d'adaptation à l'exécution supposent spécifiquement des fonctions de notation moyenne pondérée. Par contre, pour avoir une bonne optimisation dans les problèmes multicritères, une des techniques MCDA puissantes comme (WSM) adaptée peut être utile pour gérer la qualité et la scalabilité de la sélection TOP-k. L'approche proposée donne de meilleurs résultats trouvés de temps de réponse et de qualité par rapport aux autres algorithmes comme ceux qu'on trouve dans le ce travail de Fagin: US2003220921A.

En outre, les grandes infrastructures de données et les systèmes de recommandation distribués nécessitent l'utilisation d'un algorithme parallèle. Il y a cependant un intérêt croissant pour la parallélisation des requêtes TOP-k.

De plus, chaque algorithme a sa préférence par rapport aux autres algorithmes, des accès moins triés facilitent la gestion du traitement des requêtes TOP-k. Les solutions existantes sont des parallélisations utilisant des algorithmes déjà existants, et toutes les approches partagent l'idée de partitionner l'ensemble des données en sous-ensembles, de traiter les sous-ensembles localement en parallèle, et finalement de fusionner les résultats.

Aujourd'hui, à l'ère des Big Data, l'importance du traitement TOP-k est primordiale. Il nécessite une plate-forme évolutive, car il est pratiquement impossible pour les utilisateurs d'inspecter un

grand nombre de résultats de requêtes non classifiées, simplement en raison de leur volume extrême.

L'approche technique pour l'application de méthodes proposées dans un contexte distribué consiste à modifier les algorithmes sous-jacents pour utiliser le modèle de programmation spécifique à la technologie informatique utilisée.

Pourtant, malgré l'importance du problème, il y a un manque d'algorithmes entièrement parallèles qui fonctionnent efficacement à grande échelle et qui révèlent probablement les résultats corrects et exacts du cadre Map-Reduce (Saouk et coll., 2016).

Par conséquent, pour stocker de grands volumes de données, les systèmes de fichiers distribués sont une solution possible. Dans le contexte de l'analyse de données, un modèle distribué est avantageux à la fois en raison de la capacité de pousser le calcul vers différents nœuds d'un cluster et de l'évolutivité qu'il offre. Par conséquent, les cadres d'applications qui ont un code ouvert Hadoop et Spark sont devenus synonymes de Big data.

Le système de fichiers distribué couplé au moteur Map-Reduce dans le projet Hadoop d'Apache. Cependant, Apache Spark, en tant que nouvelle plate-forme d'analyse de données open source, offre une gamme de fonctionnalités beaucoup plus complète que MapReduce de Hadoop puisqu'il résout les algorithmes itératifs en utilisant la mémoire interne. Il est considéré comme exécutant des programmes beaucoup plus rapidement que son homologue Map Reduce principalement parce qu'il utilise (Resilient Distributed Datasets) RDDs comme bloc de programmation.

Aussi, l'efficacité de Spark est également due au concept de l'évaluation paresseuse. Les transformations construisent un nouveau RDD à partir d'un RDD précédent en fonction de certaines conditions. Il facilite la gestion d'algorithmes complexes et itératifs pour Spark. Par exemple, les techniques CF basées sur la Méthode SVD sont au cœur de nombreux algorithmes TOP-k. Malheureusement, Funk SVD ne peut pas être utilisé seul, car nous pouvons rencontrer un problème commun dans le moteur de recommandation appelé "Cold Start Problème" ou bien le démarrage à froid. Ce problème signifie que nous ne pouvons pas faire de recommandations pour de nouveaux utilisateurs (médecins par exemple) qui cherchent de nouveaux traitements ou services. Par contre l'ancien approche permet de sélectionner seulement traitement qui est déjà connu par ces chercheurs. Et non pas une nouvelle suggestion selon son besoin qui pourra être meilleurs que ce qui est existant. Pour remédier à ce problème, une approche réussie

consiste à combiner le modèle de Funk SVD qui se base sur les techniques de CF avec des méthodes telles qu'un algorithme basé sur le classement TOP-k, le traitement des requêtes à dominance k, et les méthodes basées sur le contenu.

L'état de l'art des solutions proposées va dans le même sens que notre démarche d'optimisation. Où nous utilisons le modèle Funk SVD avec le traitement des requêtes Skyline et TOP-k en utilisant MCDA.

Néanmoins, le calcul SVD d'une grande matrice creuse est souvent effectué avec un algorithme itératif basé sur la multiplication de cette matrice creuse. L'application de RDD pour implémenter les algorithmes présentés facilite cette tâche grâce à la bibliothèque Spark MLlib qui est une bibliothèque d'apprentissage automatique, sa naissance était avec la version 1.2 de Spark, qui contient tous les algorithmes et utilitaires d'apprentissage classiques, comme la classification, la régression, le clustering, le filtrage collaboratif (FC) et la réduction de dimensions.

BRÈVE DESCRIPTION DE L'INVENTION

L'invention a pour objet de présenter une approche de recommandation présentée comme dans la figure 1. En effet cette approche est présentée dans le cadre d'un système qui se base sur de différents composants.

La création de systèmes de recommandation hybrides en générale implique la mise en œuvre de systèmes distincts basés sur la collaboration et le contenu. Nous pouvons alors avoir deux scénarios différents :

Soit en combinant les résultats obtenus à partir des différents systèmes de recommandation en une seule recommandation finale. Ou bien en utilisant l'une des recommandations individuelles à tout moment, en choisissant celle qui est la plus appropriée que les autres sur la base d'une mesure de la qualité de la recommandation.

Dans la présente invention, l'approche proposée distingue deux scénarios différents.

Toutefois, elle utilise un système de recommandation générique composé de deux sous-systèmes illustrés dans la figure 1. Chaque sous-système étant considéré comme un système de

recommandation hybride basé sur les méthodes MCDA et FCs. Le but du premier est le raffinement de la méthode Skyline et d'ELECTRE IS (méthode de surclassement MCDA). Le raffinement utilise le traitement des requêtes de dominance s'appliquant par TOP-k, et les techniques CFs s'appliquant par un modèle qui s'appelle Funk SVD. Il combine toutefois les avantages des algorithmes Skyline et TOP-k dans le premier sous-système commun appelé Système de recommandation numéro 1 (RS1).

BRÈVE DESCRIPTION DES DESSINS

- **La figure 1** présente le principal objet du système général de recommandation avec ses sous-systèmes RS1 (021) et RS2 (022) dans le contexte distribué dans une BD distribuée.
- **La figure 2** présente le déroulement de traitement proposé pour l'algorithme SPTopkws dans un contexte distribué, et particulièrement une extension plus détaillée de bloc 2 de Figure 1.

DESCRIPTION DÉTAILLÉE DE L'INVENTION

La figure 1 présente le principal objet du système général de recommandation. RS1 (021) est la gestion des problèmes émanant de TOP-k et de Skyline. À savoir, le problème de TOP-k utilisé sur de grandes dimensions de données, qui crée une sorte de requête coûteuse TOP-k, et le problème émanant de Skyline utilisé sur de grandes dimensions.

De plus, le Pre-Skyline Processing Agent (PSPA) (031) gère l'opérateur Skyline et l'algorithme ELECTRE IS selon les meilleures exigences de l'algorithme TOP-k. Les services cloud renvoyés et leurs dimensions sont stockés sous forme de données échantillonnées. Ce système de recommandation hybride optimise la taille du résultat et le temps de réponse de la requête tout en maintenant l'efficacité et la précision des résultats obtenus par l'Agent de Recherche et de Sélection (RSA1)(032). Le degré de maximisation ou de minimisation des critères donnés est fixé par l'utilisateur final. Par exemple, les critères qui doivent être minimisés dans un service cloud sont la disponibilité, la perte de données, la latence, le coût permanent, le risque, la mémoire vive et le temps de réponse. De plus, les critères qui doivent être maximisés sont la Bande passante, le Disque dur et la Portabilité.

Comme mentionné dans la Figure 1. Le RS1 qui est basé sur (03) et (05) gère les inconvénients de Skyline et les inconvénients de TOP-k tout en utilisant les principaux avantages des paradigmes combinés. Malgré cela, le RS1 n'a pas couvert tous les scénarios de recommandation possibles, car il existe encore certains scénarios dans lesquels l'utilisateur peut déterminer la fonction de classement, où l'étape Skyline présentée par le composant (03) n'est pas nécessaire.

Par conséquent, l'élimination de l'étape Skylin en peut entraîner une réduction de la durée d'exécution de la recommandation. Cette démarche nous incite à généraliser le Système tout en offrant une solution de système complet et avec plus d'équilibre. Le nouveau système est appelé Système Générique de Recherche et de Sélection (SGRS) en ajoutant un Système de recommandation numéro 2 (RS2) figurant dans (022). Par la suite, le RS2 est utilisé pour les requêtes dominantes de TOP-k qui utilisent directement l'algorithme SPTOP-k WS (051) qui présente l'algorithme TOP-k que nous avons conçu et (052) qui présente WS (la méthode de somme pondérée utilisée dans la fonction score de cet algorithme comme une sorte d'hybridation entre deux approches. En plus, l'algorithme TOP-k est appelé TOP-k WS dans le cas d'une base de données centralisée et SPTOP-k WS dans le cas d'une base de données distribuée comme indiqué dans la Figure 1.

Dans cette figure on décrit aussi les composants principaux de ces deux sous-systèmes RS1 et RS2. L'interface utilisateur bloc 1 (011) permet aux utilisateurs RS d'interagir avec le système en sélectionnant les exigences auxquelles les services doivent répondre et de visualiser les résultats renvoyés. Elle permet également aux utilisateurs d'ajouter des éléments en renseignant leurs attributs.

Nous estimons que ces exigences sont le point commun aux ontologies des services existantes (selon le cas de chaque domaine d'application par exemple le Cloud Computing, les films, le meilleur joueur, etc.) dans le cadre du domaine d'application.

L'agent de traitement du prétraitement Skyline (PSPA) (031) prépare les résultats extraits de la base de données par le RSS (032) pour l'exécution de l'opérateur de PSPA. Les éléments retournés et leurs dimensions sont stockés sous forme des échantillons. Les dimensions utilisées sont les critères qui présentent le choix donné par l'utilisateur de système qui ne sont pas "fixe", et donc doivent être optimisés, par exemple dans le cas de critères d'un service cloud

Computing avec le prix (à minimiser), la largeur de bande (à maximiser), la latence du réseau (à réduire au minimum).

Enfin, le traitement du GRSS et plus précisément les deux sous-systèmes RS1 et RS2 comprennent l'agent de traitement des requêtes des utilisateurs (UQPA) (02) en commun.

Le traitement des demandes appliquées a été intégré dans l'utilisation d'une étape ultérieure en employant la technique de MCDA la plus connue présentée par la méthode de la somme pondérée (WSM) (052) et basée sur la théorie de l'utilité.

L'exploitation des sorties de Skyline se fait en deux façons grâce aux deux RS cités auparavant, ensuite ces données sont utilisées comme entrée de l'algorithme SPTopkws qui se base sur l'exploitation du modèle Funk SVD schématisé dans le bloc (06). Les données sortant de Skyline et d'ELECTRE IS (présenté par l'agent EISA (033)) est utilisé comme l'entrée de l'algorithme parallèle TOP-k dans le cas de RSA1, ou bien des données qui passent directement via l'agent de recherche et de sélection 2 (RSA2) (041) dans le cas de RSA2.

Par la suite, nous avons adopté cette méthode comme une méthode de la somme pondérée biobjective ou bien en anglais, Biobjective Weighted sum Method (BWSM), qui est la fonction de score pour l'algorithme SPTOP-kw parallèle détaillé dans figure 2. Cet algorithme est exécuté dans une base de données distribuée basée sur Spark RDD en exploitant le HDFS en anglais Hadoop Distributed File System et Spark afin d'exécuter les requêtes et générer les résultats de manière parallèle grâce au RDD et un mémoire distribué. Le cluster sur lequel se base l'architecture de Big data est situé ans le Cloud tel que l'architecture consiste à construire un total de trois machines virtuelles dans le cluster Apache Spark : l'une est le nœud principal (maître), les autres sont les esclaves (travailleurs).

D'autres caractéristiques et avantages de l'invention vont être détaillés ci-après.

Pour détailler le fonctionnement de l'algorithme Sptopkws dans l'ensemble de systèmes qui est représenté dans la figure1 Bloc2, on propose aussi la figure2 qui décrit les 6 étapes principales de ce traitement.

Etape1 : introduction du poids en fonction de chaque utilisateur. Un nouvel utilisateur saisit ses éléments préférés, pour le modèle.

Pour être précis, cette étape c'est une phase préparatoire pour la construction de la matrice grâce à la méthode de factorisation matricielle qui est une factorisation d'une matrice en un produit de matrices. Dans ce cas du filtrage collaboratif, les algorithmes de factorisation matricielle

fonctionnent en décomposant la matrice d'interaction de l'utilisateur du service de deux matrices rectangulaires de dimension inférieure. Une matrice peut être considérée comme la matrice de l'utilisateur où les lignes représentent les utilisateurs et les colonnes sont des facteurs latents. L'autre matrice est la matrice de l'article où les lignes sont des facteurs latents et les colonnes représentent les articles.

Etape2 : Collecte d'informations dans la matrice des éléments de l'utilisateur qui sont les nouveaux échantillons d'interaction utilisateur-élément. Le nombre de facteurs latents peut être réglé par une validation croisée ce qui est le cas de la présente invention. Les facteurs latents sont les caractéristiques de l'espace latent de dimension inférieure projetées à partir de la matrice des interactions entre l'utilisateur et l'objet.

Etape3 : Utilisation de factorisation matricielle basée sur Funk SVD modèle de la bibliothèque MLIB de Spark d'une manière parallèle.

Les poids utilisés pour calculer la fonction score pondérée dans l'Étape4 (sont basés sur la similarité en utilisant cette méthode de Funk SVD, par contre les poids qui sont calculés dans la partie fonction score de SPTopkws (WSMA) sont basé sur un scénario de choix donné par l'utilisateur final de système voir Étape5 figure 2.

Étape4 : Le système fait des prédictions de notation basées sur le modèle SVD sur tous les éléments pour cet utilisateur.

Étape5 : calculer les recommandations k plus importantes pour chaque utilisateur sur la base de la fonction de score adaptée à l'aide de SPTopkws.

L'objectif de SPTopkws se base sur le calcul des éléments les plus pertinents basé sur les critères pondérés. La WSM adaptée dans l'équation1 qui est la fonction score de cet algorithme a pour but de refléter correctement la préférence de l'utilisateur final (le médecin de l'exemple précédent à qui cherche un nouveau traitement à titre d'exemple) avec une agrégation de deux équations monotones et linéaires en une seule fonction score, voir l'équation1.

$$\begin{cases} \text{Max} \sum (w_j a_{ij}) \\ \text{Max} \sum \left(w_j \times \left(\frac{1}{a_{ij}} \right) \right) \end{cases}$$

Equation 1: $\text{Max} \sum (w_j a_{ij}) + \text{Max} \sum \left(w_j \times \left(\frac{1}{a_{ij}} \right) \right)$

Tel que les poids $\sum_{i=1}^p \omega_i = 1$, $\omega_i \geq 0$, ω_i , $i \in \{1, \dots, p\}$ sachant que les a_{ij} sont les critères de produit ou service concerné par exemple le traitement demandé.

Étape6 : Afficher les recommandations à l'utilisateur

APPLICATION INDUSTRIELLE

La présente invention a été appliquée dans différents domaines d'application, à savoir la sélection des services cloud, la sélection des meilleurs joueurs selon la base de données de la FIFA et enfin la sélection des meilleurs films selon la base de données MoviLen etc. Ce type de traitement peut aussi être appliqué dans l'aide à la prise de décision dans le domaine de santé, en exploitant la prédiction dans le diagnostic des maladies ou de trouvé un traitement approprié parmi une Base de données gigantesque existante en un temps réel et de manière distribuée. En outre, ce système n'est nullement limité aux modes de production décrits et représentés, mais le professionnel qualifié pourra en faire toute variante conformément à son concept.

REVENDEICATIONS

1. Un système de recommandation et d'aide à la décision à base de calcul de la fonction score TOP-k composé de trois unités :
 - Une première unité assurant l'interaction entre l'utilisateur et le système composée de :
 - a. Une interface pour lancer des requêtes multicritères.
 - b. Un 'agent de traitement des requêtes de l'utilisateur (UQPA) (02).
 - c. Une première entrée du système de recommandation numéro 1 (RS1) (021).
 - d. Une seconde entrée de système de recommandation numéro 2 (RS2) (022).
 - Une seconde unité assurant le prétraitement des données provenant de ladite première unité composée d'un :
 - a. Agent de prétraitement de Skyline (PSPA) (031)
 - b. Agent (RSA1) de recherche et de sélection (032)
 - c. Agent ELECTRE IS (ESISA) (033)
 - d. Agent (RSA2) de recherche et de sélection (041).
 - Une troisième unité (05) composée de :
 - a. Une base de données distribuée (07).
 - b. Un agent de traitement de requête Topk (TopkQPA) (051).
 - c. Un agent de la méthode de somme pondéré (052) qui calcule la fonction score de TOP-k.
 - d. Un agent prédicteur de classement basé sur Funk SVD (06)
2. Le système selon la revendication 1 caractérisé en ce que ladite unité 3 effectue des traitements selon un procédé comprenant les étapes suivantes :

Étape1 : reçoit les données prétraitées de ladite unité2.

Étape2 : Construit la matrice des éléments d'interaction utilisateur-élément

Étape3 : Utilise la FM basée sur Funk SVD en Spark MLIB

Étape4 : Établit les prédictions

Étape5 : Calcul les TOP-k en utilisant l'algorithme SPTopkws

Étape6 : Affiche les recommandations

3. Le système selon les revendications 1 et 2 caractérisé en ce que dans le cas où la fonction score de l'utilisateur est bien déterminée ladite unité2 exécute un traitement en passant uniquement par l'agent RSA2 (04) qui assure une interaction directe entre ladite unité2 et ladite unité3 sans intermédiaire.
4. Le système selon la revendication 1 caractérisé en ce que ledit algorithme Topk parallèle est exécuté en deux scénarios suivant :
 - Soit il utilise les entrées provenant de Skyline(03).
 - Soit il utilise directement la requête donnée par l'utilisateur passant par RSA2 (04).
5. Le système selon les revendications précédentes caractérisé en ce que le dit algorithme Topk parallèle SPTOPKWS Construit la matrice d'interaction basée sur utilisateur-élément en faisant appel à la bibliothèque Spark MLIB pour exécuter les agents (051) (052) et (06).
6. Le système selon la revendication 1 caractérisé en ce que ledit algorithme SPTOPKWS s'exécute en parallèle en utilisant les données distribuée (08) située dans le Cloud ; ledit algorithme exécute le code qui se base sur la fonction score BWSM ainsi que les autres fonctions itératif de l'algorithme, qui se présente par Query1 Query2 et Query3. etc. (10) en exploitant une mémoire distribuée (9) moyennant le RDD de Spark pour finalement afficher les résultats les plus pertinents (11) de manière parallèle.
- 7.

Dessins :

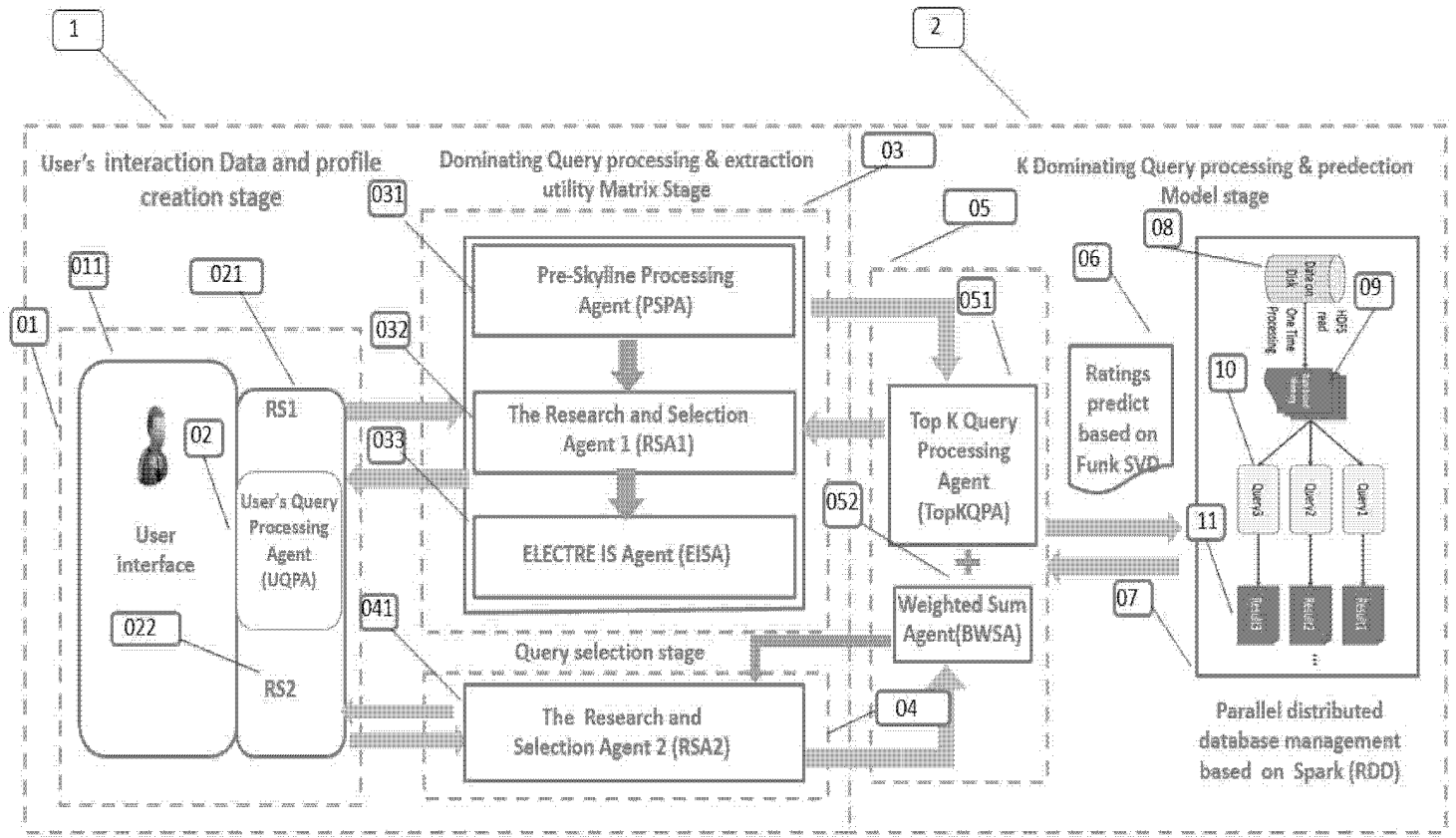


Figure 1

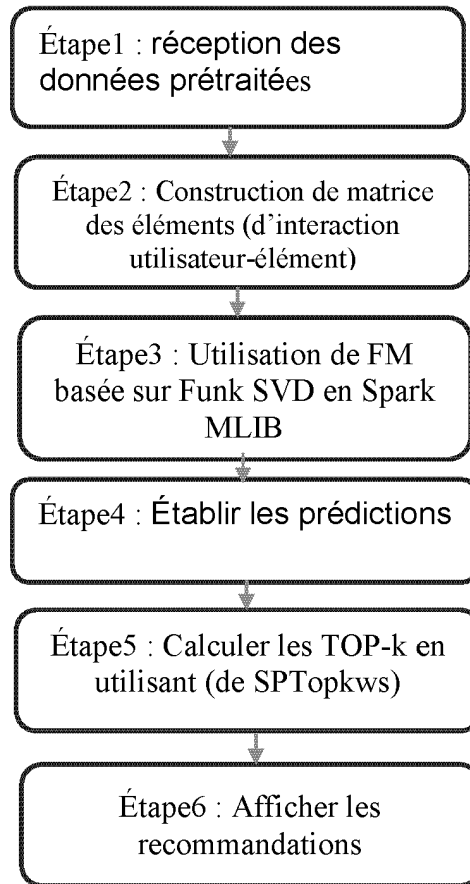


Figure 2

**RAPPORT DE RECHERCHE
AVEC OPINION SUR LA BREVETABILITE**
(Conformément aux articles 43 et 43.2 de la loi 17-97 relative à la
protection de la propriété industrielle telle que modifiée et complétée
par la loi 23-13)

Renseignements relatifs à la demande	
N° de la demande : 50776	Date de dépôt : 03/09/2020
Déposant : Université Mohammed V de Rabat	
Intitulé de l'invention : System collaboratif d'aide à la décision à base des recommandations multi critères	
Le présent document est le rapport de recherche avec opinion sur la brevetabilité établi par l'OMPIC conformément aux articles 43 et 43.2, et notifié au déposant conformément à l'article 43.1 de la loi 17-97 relative à la protection de la propriété industrielle telle que modifiée et complétée par la loi 23-13.	
Les documents brevets cités dans le rapport de recherche sont téléchargeables à partir du site http://worldwide.espacenet.com , et les documents non brevets sont joints au présent document, s'il y en a lieu.	
Le présent rapport contient des indications relatives aux éléments suivants :	
Partie 1 : Considérations générales	
<input checked="" type="checkbox"/> Cadre 1 : Base du présent rapport	
<input type="checkbox"/> Cadre 2 : Priorité	
<input type="checkbox"/> Cadre 3 : Titre et/ou Abrégé tel qu'ils sont définitivement arrêtés	
Partie 2 : Rapport de recherche	
Partie 3 : Opinion sur la brevetabilité	
<input type="checkbox"/> Cadre 4 : Remarques de forme et de clarté	
<input type="checkbox"/> Cadre 5 : Défaut d'unité d'invention	
<input type="checkbox"/> Cadre 6 : Observations à propos de certaines revendications exclues de la brevetabilité	
<input checked="" type="checkbox"/> Cadre 7 : Déclaration motivée quant à la Nouveauté, l'Activité Inventive et l'Application Industrielle	
Examineur: Sara AGUENDICH	Date d'établissement du rapport : 09/11/2020
Téléphone: 212 5 22 58 64 14/00	



Partie 1 : Considérations générales**Cadre 1 : base du présent rapport**

Les pièces suivantes de la demande servent de base à l'établissement du présent rapport :

- Description
11 Pages
- Revendications
6
- Planches de dessin
2 Pages

Partie 2 : Rapport de recherche

Classement de l'objet de la demande :

CIB : G 06F 16/00 ; G 06F 17/00 ; G 06F 17/30

CPC : G 06F 16/283 ; G 06F 16/245 ; G 06F 17/00 ; G 06F 17/30

Plateformes et bases de données électroniques de recherche :

EPOQUENET, WPI, ScienceDirect, IEEE, ORBIT

Catégorie*	Documents cités avec, le cas échéant, l'indication des passages pertinents	N° des revendications visées
A	A Survey of Top-k Query Processing Techniques in Relational Database Systems; Ihab F Ilyas, George Beskales, Mohamed A Soliman ; 15-10-2008	1-6
A	KR101271277B1 ; POSTECH ACAD IND FOUND [KR] ; 07-06-2013 <i>Abrégé ; Description</i>	1-6
A	KR100994724B1 ; POSTECH ACAD IND FOUND [KR]; 16-11-2010 <i>Abrégé ; Description</i>	1-6
A	CN106055674A ; UNIV SOUTHEAST ; 26-10-2016 <i>Abrégé ; Description</i>	1-6

***Catégories spéciales de documents cités :**

-« X » document particulièrement pertinent ; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément
-« Y » document particulièrement pertinent ; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier
-« A » document définissant l'état général de la technique, non considéré comme particulièrement pertinent
-« P » documents intercalaires ; Les documents dont la date de publication est située entre la date de dépôt de la demande examinée et la date de priorité revendiquée ou la priorité la plus ancienne s'il y en a plusieurs
-« E » Éventuelles demandes de brevet interférentes. Tout document de brevet ayant une date de dépôt ou de priorité antérieure à la date de dépôt de la demande faisant l'objet de la recherche (et non à la date de priorité), mais publié postérieurement à cette date et dont le contenu constituerait un état de la technique pertinent pour la nouveauté

Partie 3 : Opinion sur la brevetabilité**Cadre 7 : Déclaration motivée quant à la Nouveauté, l'Activité Inventive et l'Application Industrielle**

Nouveauté	Revendications 1-6	Oui
	Revendications aucune	Non
Activité inventive	Revendications 1-6	Oui
	Revendications aucune	Non
Application Industrielle	Revendications 1-6	Oui
	Revendications aucune	Non

Il est fait référence aux documents suivants. Les numéros d'ordre qui leur sont attribués ci-après seront utilisés dans toute la suite de la procédure

D1 : A Survey of Top-k Query Processing Techniques in Relational Database Systems

1. Nouveauté

Aucun des documents cités ci-dessus, considéré isolément, ne divulgue un système de recommandation et d'aide à la décision à base de calcul de la fonction score TOP-k comprenant l'ensemble des caractéristiques techniques énoncées dans la revendication indépendante 1. D'où l'objet de ladite revendication est nouveau au sens de l'art. 26 de la loi 17-97 telle que modifiée et complétée par la loi 23-13. Par conséquent, les revendications dépendantes 2 à 6 sont aussi nouvelles.

2. Activité inventive

Le document D1, qui est considéré comme l'état de la technique le plus proche de l'objet de la revendication 1, divulgue :

Un système de recommandation et d'aide à la décision à base de calcul de la fonction score TOP-k composé de :

- a) Une interface pour lancer des requêtes multicritères.
- b) Un agent de traitement des requêtes de l'utilisateur (UQPA)
- c) Agent de prétraitement de Skyline (PSPA)
- d) Agent (RSA) de recherche et de sélection
- e) Une base de données distribuée
- f) Un agent de traitement de requête Top-k (TopkQPA)
- g) Un agent de la méthode de somme pondéré qui calcule la fonction score de TOP-k.

Par conséquent, l'objet de la revendication 1 diffère de D1 en ce que le système de recommandation comprend :

- a) Une première entrée du système de recommandation numéro 1 (RS1)
- b) Une seconde entrée du système de recommandation numéro 2 (RS2)
- c) Un agent ELECTRE IS (ESISA)
- d) Un autre agent (RSA2) de recherche et de sélection

e) Un agent prédicteur de classement basé sur Funk SVD.

L'effet technique de ces différences est celui d'optimiser la taille du résultat et le temps de réponse de la requête et améliorer l'efficacité de la méthode de somme pondérée.

Le problème objectif technique que la présente invention se propose de résoudre est celui d'améliorer les processus de recommandation dans les bases de données distribuées et gérer les problèmes émanant des requêtes TOP-k et Skyline.

La solution à ce problème proposée dans la revendication 1 n'est pas décrite dans l'art antérieur, pris seul ou en combinaison. Aucun enseignement n'a été trouvé dans les documents de l'état de la technique qui aurait incité l'homme du métier, d'arriver à la solution telle que décrite dans la revendication 1.

Par conséquent, l'objet de la revendication 1 implique une activité inventive au sens de l'article 28 de la loi 17-97 telle que modifiée et complétée par la loi 23-13 concernant l'activité inventive.

Les revendications 2 à 6 dépendent de la revendication 1 dont l'objet est considéré inventif, comme indiqué auparavant, et satisfont donc aux exigences de l'article 28 de la loi 17-97 telle que modifiée et complétée par la loi 23-13.

3. Application industrielle

L'objet de la présente invention est susceptible d'application industrielle au sens de l'article 29 de la loi 17-97 telle que modifiée et complétée par la loi 23-13, parce qu'il présente une utilité déterminée, probante et crédible.