

Intégration des techniques du Datamining et des bases de données avancées dans le processus de Gestion des Connaissances : proposition d'un processus hybride basé sur le raisonnement à partir de cas

MEKROUD Nouredine¹, MOUSSAOUI Abdelouahab², SLIMANI Yacine³

Faculté des Sciences de l'Ingénieur, Université Ferhat Abbas, Sétif

¹mekroud_n@yahoo.fr,

²moussaoui.abdel@gmail.com,

³slimani_y09@univ-setif.dz

Résumé : Réutiliser les expériences passées dans la résolution des nouveaux problèmes est fréquent dans la vie quotidienne de chacun ; puisqu'il est bien évident et légitime que les problèmes similaires auront des solutions similaires, et qu'on se trouve souvent face à un problème que l'on a déjà rencontré. Le raisonnement à partir de cas (RàPC), comme une méthodologie puissante de conception des systèmes intelligents, peut être renforcé dans les différentes étapes de son processus par les techniques du Datamining et les solutions extraites du domaine des bases de données avancées. Cette hybridation fournira au RàPC les solutions techniques nécessaires pour l'implémentation des systèmes de retour d'expérience conçus. Cet article présentera un état de l'art sur les travaux d'intégration des techniques du Datamining dans le cycle du RàPC. Notre contribution sera par la proposition d'un processus hybride RàPC-Datamining, renforcé par des techniques d'optimisation empruntées des VLDB (Very Large DataBases). Des réflexes et arguments pertinents accompagneront la présentation du processus.

Mots Clés : intelligence artificielle, gestion des connaissances, raisonnement à partir de cas, Datamining, bases de données avancées.

Introduction

L'application du Retour d'Expérience est fréquente dans la vie quotidienne de chacun. Les expériences passées capitalisées représentent un patrimoine immatériel pour les individus, les compagnies et même les nations. Actuellement, la valorisation économique des entreprises repose de plus en plus sur des facteurs immatériels, notamment celui de la connaissance. La valorisation du patrimoine de connaissances (Knowledge Capital) des entreprises modernes correspond à 217% de leur capital financier net [8]. La capitalisation du savoir faire et des meilleures pratiques d'une entreprise nécessite la mise en place d'un système permettant de fournir à une personne, pas nécessairement hautement qualifiée, l'information utile au moment où elle en a besoin, de façon exploitable pour superviser les décisions à prendre.

Les connaissances d'une entreprise sont considérées comme un patrimoine fragile, et peuvent disparaître par un départ en retraite, une mutation ou un licenciement d'un expert. Cette richesse d'expertise doit être donc : capturée, capitalisée, protégée et distribuée. Cela est possible par la création des mémoires d'entreprise, ce qui aidera à rapprocher le niveau de connaissances individuel au niveau d'expérience collective de l'entreprise.

Le RàPC est une méthodologie puissante de conception des systèmes intelligents ; elle est basée sur la réutilisation des solutions des cas similaires passées pour la résolution des nouveaux problèmes rencontrés, cela en capitalisant les connaissances acquises des expériences passées. Cette méthodologie enrichit en contenu la mémoire d'entreprise par un apprentissage basé sur les expériences vécues. Ainsi, le RàPC est une méthodologie et pas une technologie de résolution des problèmes; la gamme des solutions techniques utilisées dans la réalisation de son processus est vaste.

Le Datamining, utilisé dans la découverte et la modélisation des informations utiles cachées dans une masse de données grande et complexe, offre des solutions techniques incontournables dans la découverte et la capitalisation des connaissances de l'entreprise. Aussi, vu la structuration des bases de cas des systèmes RàPC, qui est comparable à celles des bases de données, les solutions diverses de visualisation et d'optimisation des VLDB, issues du domaine des bases de données avancées, peuvent améliorer la performance des systèmes RàPC réalisés. L'objectif de ce travail est d'étudier l'utilisation des techniques du Datamining et des bases de données avancées, qui émergent actuellement, au cours du cycle RàPC ; en éclairant l'appui qu'elles peuvent fournir pour améliorer la fiabilité de cette méthodologie.

La section 2 propose une présentation générale de la gestion des connaissances. La section 3 étudie la méthodologie du raisonnement à partir de cas. La section 4 présente l'état de l'art des travaux réalisés concernant l'intégration des techniques du Datamining, et des bases de données avancées, dans les différentes étapes du cycle RàPC. Dans la section 5, on propose une démarche intégrant des algorithmes du Datamining et des VLDB dans le processus du RàPC. On finira par une conclusion et des perspectives.

1. Les connaissances dans l'entreprise

1.1. Présentation générale

L'un des aspects de développement de l'efficacité d'une entreprise est la capitalisation de ses expériences, par la création d'une plateforme d'expertise donnant la possibilité de réutilisation du savoir et savoir-faire et des compétences développées au cours du temps, vu le risque de centralisation du savoir faire dans les experts humains qui peut causer la non disponibilité de cette richesse suite à une sur-occupation de l'expert, mutation, départ à la retraite, démission, licenciement etc. [6].

Les connaissances sont une représentation réduite du monde réel [7]. Dans une entreprise, deux catégories de connaissances sont distinguées, à savoir : les connaissances Tacites (implicites, non formalisables) : qui sont difficiles à décrire, comme les compétences, les habilités, la connaissance historique de l'organisation ...etc. et les connaissances Explicites (formalisables) : qui sont codifiables plus facilement (manuels, plans, modèles, documents d'analyse, données, ...). Par exemple : avant l'invention du système de notation (solfège) au XII^{ème} siècle, on ne pouvait pas représenter les connaissances musicales ; elles pouvaient uniquement être apprises par l'expérience directe, l'écoute (tacite) ; mais grâce au solfège, ces connaissances sont devenues explicites [4].

Les éléments de la connaissance sont [16] :

La donnée (quoi) : c'est un élément de base (mesure ou caractéristique) pour la représentation de l'information dans les bases de données (ex : 100°C).

L'information (qui, quand, où) : est une donnée contextualisée qui représente un fait réel (ex: l'eau bout à 100°C).

La connaissance (comment, pourquoi) : C'est l'information interprétée ayant une signification (ex : lorsque l'eau bout à 100°C en utilisant une source de chaleur, les microbes seront éliminés).

La compétence (action et réutilisation) : c'est une connaissance exploitable pour accomplir une tâche ou une action (ex : maîtriser les conditions d'ébullition de l'eau pour avoir un liquide stérilisé).

On doit gérer deux types de connaissances dans l'entreprise, le savoir et le savoir faire [4] :

Le savoir : ce qui représente les compétences individuelles et collectives, connaissances des faits et des dépendances, plans, documents, modèles, données, etc. Le savoir a le caractère des connaissances explicites et correspond plutôt à la partie gestion de l'information représentée par des objets.

Le savoir-faire : représente la connaissance de la résolution des problèmes, procédures, stratégies, contrôle des actions, expériences individuelles et de l'organisation ; il a plutôt le caractère tacite. Cela correspond à la gestion des hommes et est représenté par des processus.

1.2. La gestion des connaissances

Dans la littérature de la gestion moderne des entreprises, on remarque l'évocation de nouvelles notions comme : les travailleurs de la connaissance, la société de la connaissance, la mémoire d'entreprise, l'apprentissage et la cognition organisationnelle, les entreprises apprenantes, la gestion des richesses immatérielles, la gestion de l'innovation...etc., où la richesse de l'entreprise est désormais basée sur ses activités intellectuelles, donnant naissance à une économie basée sur la connaissance [7].

La gestion des connaissances, ou Knowledge Management en anglais (baptisée KM), désigne un ensemble de concepts et d'outils permettant la production des connaissances et le développement des compétences individuelles, collectives et organisationnelles. Le KM vise à rassembler le savoir et le savoir faire sur des supports accessibles, et faciliter leur transmission en temps réel à l'intérieur de l'établissement, ou les différer à nos successeurs [16]. C'est une application pratique des sciences cognitives, de l'intelligence artificielle et des sciences de l'organisation [15].

Le cycle de capitalisation des connaissances est composé de quatre phases : la première consiste à « repérer » les connaissances cruciales de l'entreprise ; Ensuite, il faut les « préserver », c'est-à-dire les modéliser, les formaliser et les conserver ; Puis, les « valoriser » et les mettre en exploitation en les rendant accessibles ; Après, les « actualiser » et les enrichir progressivement en les gérant continuellement [15].

Plusieurs cycles drivés de capitalisation des connaissances peuvent être rencontrés dans la littérature [6] : Lai & chu (2000) proposent un cycle constitué des étapes suivantes : initiation, élaboration, modélisation, stockage, distribution, utilisation et révision. Pour Andersen (1996) : c'est identifier, collecter, créer, organiser, partager, appliquer et adapter. Grundstein (2000) : détecter (identifier et localiser), préserver (modéliser, formaliser, stocker), capitaliser (accès, diffuser, utiliser, intégrer), actualiser (mise à jour, enrichir). Tannembaum & alliger (2000) : partage, accès, assimilation, application ... etc.

Aussi, il existe une diversité de méthodologies de capitalisation des connaissances d'une entreprise, celles spécifiques à la construction des mémoires d'entreprises, à savoir : REX, MEREX, CYGMA, atelier FX et Componential Framework ... etc, et d'autres empruntées de l'ingénierie des connaissances, qui peuvent être subdivisées en deux approches : ascendante et descendante [16]. L'approche descendante sépare la connaissance du domaine de son utilisation, et permet ainsi de définir des composants génériques réutilisables (ex : KADS, CommonKADS, MKSM, MASK...) [6] ; tandis que l'approche ascendante présente un modèle structuré d'expertise, certes, mais avec une abstraction les terminologies propres au problème (ex : KOD, MÉTODAC...) [4].

2. Le raisonnement à partir de cas

Le RàPC (en anglais Case-Based Reasoning) est un processus qui vise la réutilisation des expériences passées. Cette technique provenant du domaine de l'Intelligence Artificielle, a été utilisée dans les systèmes experts et les sciences cognitives. Dans cette approche, l'utilisateur essaie de résoudre un nouveau problème en reconnaissant les similarités avec des problèmes préalablement résolus, appelés cas. Un cas est communément un problème spécifique qui a été identifié, résolu, stocké et indexé dans une mémoire avec sa solution, et éventuellement le processus d'obtention de celle-ci [4]. Les systèmes de RàPC sont appliqués dans de nombreux domaines comme : la médecine, le commerce, les services de consultation, la maintenance, le contrôle, l'analyse financière.

2.1. Le cycle RàPC

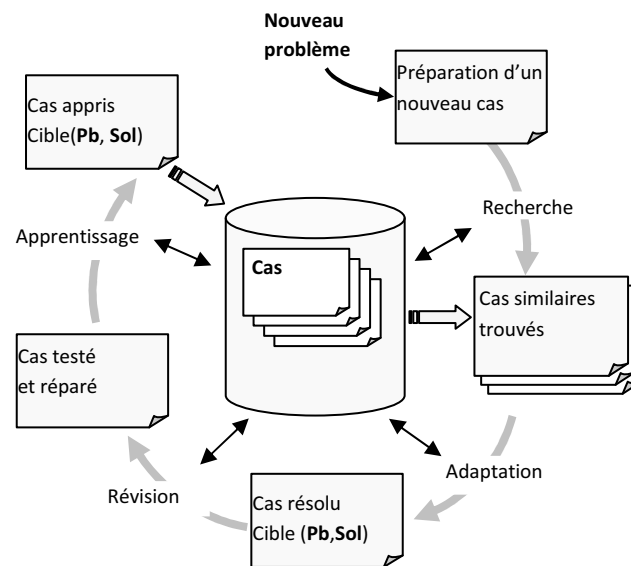


Figure 1 : Le cycle du raisonnement à partir de cas [15]

Les étapes de ce cycle peuvent être présentées comme suit [16] :

1. L'élaboration d'un nouveau problème (cas cible) : représente l'acquisition des informations connues sur le nouveau problème, pour lui donner une description initiale. Elle inclut : la structuration, la modélisation et la représentation du cas ; ce dernier doit être représenté d'une manière similaire à un cas source.
2. La remémoration des cas (sources) : rechercher les cas les plus similaires, signifie la recherche des correspondances entre les descripteurs des cas de la base et ceux du cas à résoudre. Différentes techniques peuvent être utilisées comme : le calcul d'un degré d'appariement des descripteurs (similarité entre deux cas), la pondération éventuelle des descripteurs ...etc.
3. L'adaptation des cas : réutiliser totalement ou partiellement la solution du cas sélectionné pour résoudre le nouveau problème.
4. La révision de la solution proposée : signifie l'évaluation de la solution dans le monde réel (bilan d'un cas), ainsi que la vérification par introspection de la base de cas en considérant la qualité des cas.

5. La mémorisation d'un nouveau cas : représente l'ajout éventuel du cas cible dans la base de cas. Dans cette phase on peut également synthétiser, modifier et apprendre de nouvelles connaissances.

2.2. Avantages et inconvénients

L'approche RàPC offre de nombreux avantages, c'est une démarche plus simple à mettre en oeuvre que celles basées sur un modèle du domaine, puisqu'elle permet d'éviter les difficultés de modélisation du savoir-faire des experts (complexité des ontologies et des représentations logique...) ; Aussi, c'est une bonne solution pour les domaines n'exigeant pas de solutions optimales, ou dont les principes sont mal formalisés ou peu éprouvés [16]. Donc, le RàPC diffère des autres approches de l'IA qui visent la représentation des connaissances générales utilisées dans un domaine sous forme d'un système expert, le RàPC vise à utiliser la connaissance spécifique et pragmatique des problèmes précédemment expérimentés, d'une façon progressive avec le temps, et l'apprentissage sera ainsi incrémental et basé sur les expériences vécues [4].

Mais, par contre, le RàPC ne trouve pas nécessairement la solution concrète à un problème ; et parfois, juste générer un ensemble de solutions possibles [5]. Aussi, vu la nécessité d'une intervention et une mobilisation en continu des experts, lors de la capitalisation progressive des connaissances, l'individualisme constitue un frein redoutable ; Les experts hésitent de partager leurs connaissances acquises après des années de travail, par méfiance d'une restructuration ou compression dans l'entreprise, ou ils estiment leurs savoir-faire comme leur plus grande assurance [6].

Cet article est basé sur l'idée que le RàPC est une méthodologie et pas une technologie de résolution de problèmes ; puisque la majorité des définitions du RàPC se basent sur la présentation du « Quoi » et pas le « Comment », et visent surtout l'explication des étapes du cycle de RàPC, ce qui est plus proche de la description d'une méthodologie et non pas une technologie particulière. L'application de cette méthodologie peut appeler à l'utilisation de différentes technologies comme celles empruntées du Datamining (le voisin le plus proche, l'induction indexée de l'apprentissage automatique, la logique floue ...). Ainsi, présenter le RàPC comme une méthodologie est un avantage primaire, et est important pour son développement. Si le RàPC est considéré comme une technologie, il pourrait sembler que les recherches dans ce domaine sont en grande partie achevées, mais s'il est considéré comme une méthodologie, les chercheurs auront toujours le défi d'introduire dans son cycle de diverses technologies [18].

2.3. Les expériences erronées

Les systèmes de résolution des problèmes à base de cas raisonnent et apprennent des expériences passées, pour guider le futur raisonnement. Les avantages attendus de ce processus d'apprentissage dépendent de deux types de régularité : (1) La régularité des problèmes-solutions : qui assurent que les solutions aux problèmes passés similaires sont des points de départ utiles pour résoudre des problèmes actuels. (2) La régularité de distribution des problèmes : qui assure que la base des cas va contenir des cas similaires aux nouveaux problèmes rencontrés. Mais, ces régularités ne sont pas toujours assurées, même si la régularité initiale est suffisante, les contraintes internes et l'environnement extérieur changent au fil du temps.

Leake et Al (1999) ont essayé de définir des normes standard pour mesurer les deux types de régularité, et proposent une approche de détection des changements nécessitant la maintenance de la base de cas, basée sur l'analyse des performances au fil du temps, pour répondre aux changements de l'environnement. Ces auteurs sont parmi les premiers qui ont signalé la possibilité d'avoir des expériences erronées, en étudiant les écarts par rapport au niveau de régularité désiré. Donc, puisque les systèmes du RàPC sont utilisés à longs termes, le travail de ces auteurs fournit un point de départ pratique pour savoir comment détecter et répondre à des situations dans lesquelles la réutilisation des expériences deviendra une erreur [11].

3. Le RàPC & le Datamining

Le Datamining est l'extraction des informations utiles, et la découverte des connaissances et des modèles cachés dans des bases de données hétérogènes. C'est un processus itératif et non-trivial de découverte d'informations implicites, précédemment inconnues et potentiellement utiles [9]. Le Datamining s'adresse à la phase d'extraction et découverte de connaissances à partir de données, en utilisant le savoir-faire de plusieurs domaines [13] : l'intelligence artificielle, l'apprentissage, la reconnaissance de modèles, l'acquisition de connaissances pour les systèmes experts, la visualisation des données ...etc.

3.1. Les techniques du Datamining et des VLDB utilisables dans le cycle du RàPC

Dans chaque étape du RàPC, diverses techniques de Datamining et des VLDB peuvent être utilisées pour renforcer la fiabilité des tâches requises, on cite :

L'élaboration : les techniques de nettoyage de données, appliquées sur les attributs des cas stockés, assurent la justesse des connaissances fournies [5].

Remémoration : la procédure de recherche des cas similaires est implémentée par diverses mesures de similarité ; Par exemple : weighted block-city, la mesure euclidienne et de Minkowski, les approches inductives, les réseaux neurones, la logique floue et surtout les K plus proches voisins (KPPV) en utilisant des métriques de similarité superficielles d'ordre syntaxique ou profondes d'ordre sémantique [16].

Pour faciliter la remémoration, la fragmentation des grandes bases de cas localise la recherche en des fragments pertinents au problème à résoudre [5] ; Aussi, on peut indexer la base par des méthodes basées sur la similarité, ou l'apprentissage inductif [16].

Révision et adaptation : quoique ces deux étapes soient les plus difficiles à modéliser, les algorithmes génétiques et les techniques de satisfaction des contraintes peuvent être utiles pour tester la fiabilité de la solution proposée [15].

Mémorisation : par l'ajout d'un nouveau cas à la base, ou l'adaptation du cas similaire au problème à résoudre. Une synthétisation des nouvelles connaissances est possible, afin d'optimiser la base des cas. Les techniques d'apprentissage automatique sont utiles dans cette phase [16].

3.2. Le Textmining et le RàPC

La méthodologie du RàPC est basée sur la recherche et la réutilisation intégrale ou partielle d'un cas similaire au cas actuel à résoudre, sans se préoccuper des informations complémentaires qui ont été nécessaires pour résoudre l'ancien cas résolu. Le contexte de résolution d'un problème, décrivant les causes et les circonstances de ce dernier, ainsi que ses conséquences et ses influences, donne les informations nécessaires pour bien assimiler la situation réelle. Les supports documentaires (rapports techniques de résolution des problèmes passés), non exploités dans le RàPC, peuvent donner un grand appui pour la compréhension du problème à résoudre, et la proposition des cas similaires les plus pertinents [12]. Vu que la résolution des problèmes est un processus complexe, et inclut une série de situations incertaines; Seule l'identification des cas similaires est parfois insuffisante pour la résolution du problème courant. On peut compenser ce manque par l'introduction des techniques du Textmining dans le processus de résolution des problèmes. Le concept des règles d'association peut tirer des profils situations-actions pour découvrir des connaissances décisionnelles.

3.3. Quelques contributions concernant la combinaison du Datamining et le RàPC

Les techniques du datamining ont été combinées avec le RàPC dans plusieurs travaux de recherche :

Vu la lourdeur de la tâche de construction de la base des cas, Pan et Al (2007) ont présenté un algorithme de construction automatique d'une nouvelle base de cas de haute qualité, à partir d'un ensemble de cas bruts, pour améliorer la compétence du processus de RàPC. Les auteurs ont analysé deux problèmes majeurs dans les algorithmes de recherche des cas similaires ; le premier est causé par les cas bruyants tels que le cas le plus proche d'un problème à résoudre qui ne fournit pas des solutions correctes pour ce dernier ; Le deuxième problème est causé par l'inégalité de la distribution des cas, tel que les problèmes similaires qui ont des solutions différentes, à cause généralement de la mauvaise maintenance de la base de cas. Pour résoudre ces problèmes, ils ont développé un cadre théorique basé sur une analyse discriminante dans le but d'éliminer les cas bruyant, suivie d'une évaluation empirique par des expérimentations extensives des cas résultants pour minimiser la marge d'erreur, et ainsi construire une nouvelle base de cas de qualité et compétence meilleures que celles de la base originale.

Kim et Han (2001) ont proposé une nouvelle méthode d'indexation et de classification de la base des cas ; en utilisant en plus des connaissances spécifiques du cas à résoudre des connaissances additionnelles dérivées du cluster qu'appartient le cas sélectionné. Des réseaux neurones artificiels compétitifs, qui peuvent produire des clusters adaptatifs meilleurs que ceux des algorithmes de segmentation basés sur les statistiques, sont utilisés pour générer les valeurs centroïdes donnant des connaissances généralisées des clusters. Ces valeurs centroïdes seront ajoutées à la base des cas, sous forme de cas artificiels représentatifs, utilisées ensuite dans le procédé d'indexation, pour faciliter la recherche des cas les plus pertinents.

Zhuang et Al (2007) ont proposé la segmentation des bases de cas volumineuses en utilisant les cartes de Kohonen pour l'auto-organisation (SOM : Self-Organising Maps), qui appartient à la classe des outils basés sur les réseaux neurones, pour l'apprentissage non supervisé. L'algorithme de SOM est basé sur la répétition du repositionnement des enregistrements dans la carte, jusqu'à la minimisation de l'erreur dans la fonction de classification. Le but de l'intégration du SOM est de trouver des groupements d'enregistrements (cas) avec un maximum de simultitude intra-groupe, et de désimultitude inter-groupes, pour bien préparer la phase de remémoration par une bonne organisation des cas similaires stockés.

À cause des incertitudes dans la représentation des connaissances et la description des attributs et des mesures de similarité, il est parfois très difficile de trouver des cas similaires dans la base de cas. Pour faire face à ce problème, Chang et All (2008) ont proposé d'incorporer la logique floue dans le processus du RàPC, pour avoir un modèle de prévision des ventes avec plus de souplesse et de précision. Cela est basé sur la sélection des ventes antérieures les plus similaires et les plus utiles à un cas courant.

N. Arshadi et Al (2000) ont proposé une nouvelle approche pour l'adaptation compositionnelle. Ils incitent la possibilité de combiner les solutions trouvées, pour permettre l'obtention d'une solution finale plus pertinente. Les solutions similaires peuvent être combinées par le choix de la valeur moyenne de leurs attributs. Une telle approche enrichit la base par des cas généralisés plus ou moins précis.

Yang et Wu (2000) ont proposé une méthode simplifiée pour la maintenance des bases de cas volumineuses, par leur segmentation en plusieurs bases plus petites en taille. Les cas appartenant au même segment seront les plus similaires possible. Ce processus donne de nouvelles bases de cas plus faciles à explorer et à indexer ; ainsi la maintenance sera transparente et sans complexité. Les auteurs ont choisi la méthode de segmentation GDBSCAN, proposée en Datamining, basée sur l'idée que les clusters sont des régions ayant une densité de points plus élevée par rapport à leur entourage. Cette méthode surmonte les défauts des techniques de segmentation inductive, vu sa simplicité d'exécution, sans la spécification préliminaire du nombre de clusters exigé.

3.4. Le RàPC et les bases de données avancées

Les VLDB (Very Large DataBase) sont des bases de données contenant de très grands nombres d'enregistrements, et qui occupent de vastes espaces physiques de stockage, mesurés généralement en téra-octets ou en billions d'enregistrements. À ce niveau on parlera surtout d'entrepôts de données (Data WareHouse) ou VLDB, et pas de simples bases de données [19]. Les techniques développées pour optimiser la visualisation et l'interrogation de ces VLDB ont prouvé leur fiabilité, et offrent de grandes performances de traitement des entrepôts de données.

L'un des avantages du RàPC est sa simplicité de représentation des connaissances par une base de cas structurée sous forme d'attributs de types alphabétique, numérique, logique ... etc., donc les cas de cette base sont comparables à des enregistrements d'une base de données. Cela ouvre les possibilités d'utilisation des techniques d'optimisation des requêtes, utilisées pour les VLDB, comme

l'indexation et la fragmentation, déjà éprouvées sur les BD relationnelles et objet, pour renforcer les performances des traitements sur les bases de cas [2].

Plusieurs techniques d'indexation sont utilisées pour accélérer l'exécution des requêtes, on opte pour l'indexation basée sur la liste des valeurs, en utilisant les attributs à faible cardinalité (à valeurs redondantes) puisqu'une base de cas peut contenir des attributs ayant cette caractéristique.

La fragmentation (verticale sur les champs et horizontale sur les enregistrements) réduit le temps d'exécution des requêtes, en ne cherchant que dans les fragments de cas les plus pertinents au problème à résoudre. Dans cet article, on opte pour la fragmentation horizontale pour éviter la séparation des attributs d'un même cas. L'indexation peut être le point de départ pour une meilleure fragmentation horizontale, qui sera suivant les différentes valeurs formant la modalité de l'attribut index.

Les points communs rapprochant les traitements sur VLDB de ceux effectués sur les bases de cas, forment l'idée de base du processus proposé dans cet article.

4. Approche proposée : l'intégration des techniques de Datamining et des bases de données avancées dans le cycle du RàPC

4.1. Présentation de l'approche

On propose dans cet article une démarche hybride, où le contexte de résolution des problèmes sera conservé et utilisé pour résoudre les problèmes futurs. Les processus spéciaux, rapports d'intervention et les erreurs de traitement seront aussi stockés à côté de la solution du cas traité. Aussi la régularisation périodique de la base des cas assurera en contenu sa performance. L'optimisation de la recherche des cas similaires est renforcée par les techniques d'indexation et de segmentation horizontale de la base des cas.

L'explication de ces étapes sera comme suite :

Étape 1 : Collection des données pertinentes : assurer la disponibilité des données appropriées au contexte du cas à résoudre, par l'application des techniques de Textmining pour collecter les documents pertinents.

Étape 2 : Indexation et fragmentation de la base : indexer la base des cas par un attribut à faible cardinalité, fragmenter horizontalement ensuite cette base selon les index générés, pour réduire la recherche au segment pertinent au contexte du problème à résoudre. La répétition de cette opération à chaque résolution d'un nouveau problème n'est pas coûteuse, puisqu'elle est d'une complexité faible.

Étape 3 : recherche des cas similaires : en assurant la meilleure adéquation entre le problème à résoudre et certains cas disponibles dans la base. L'algorithme du KPPV sera appliqué sur le segment choisi.

Étape 4 : réutilisation de la solution du cas trouvé : réutiliser les solutions des cas trouvés dans le monde réel, suivant les paramètres du nouveau cas (par un recopiage direct ou une adaptation)

Étape 5 : révision et adaptation du cas utilisé: évaluer la qualité et la pertinence de la solution proposée au cas actuel, et introduire les modifications nécessaires.

Étape 6 : maintenance de la base des cas : par l'incorporation du nouveau cas (ou la perfection d'un cas ancien) dans la base. Une réindexation de la base est possible suivant son nouveau contenu.

Étape 7 : stoker le contexte de résolution du cas traité : les commentaires des experts, et les rapports sur les actions supplémentaires, doivent être conservés pour servir comme données de contexte aux cas futurs.

Régularisation périodique de la base des cas :

Pour détecter et répondre à des situations où la réutilisation des expériences similaires passées peut engendrer une erreur, on recalcule périodiquement le taux de régularité des problèmes-solutions et des problèmes-distributions pour analyser la performance de la base des cas. Les changements au fil du temps des caractéristiques des problèmes à résoudre, et de l'environnement dans lequel les solutions seront appliquées, doivent être pris en considération

4.2. Avantages et inconvénients

Ce processus tire des avantages de l'indexation pour optimiser du temps de recherche ; Mais, il exige des contraintes d'espace mémoire et la maintenance du contenu des index, suite aux changements dans la base des cas qui doivent être reportés sur ces index.

La fragmentation horizontale réduit le temps de recherche dans la base en la limitant aux fragments de cas pertinents au problème à résoudre ; aussi, permet le parallélisme des traitements sur plusieurs fragments ; Mais, la mise à jour de la base de cas peut causer une migration des cas entre les fragments causée par le changement d'appartenance d'un cas à un fragment.

L'utilisation des techniques du Datamining et des bases de données dans la méthodologie du RàPC assure sans doute une meilleure performance pour le système RàPC ; mais les traitements qu'elles exigent présenteront une charge supplémentaire aux ingénieurs de connaissances, malgré que cet engagement est rentable aux moyens et longs termes.

Le schéma suivant illustre les étapes de ce processus :

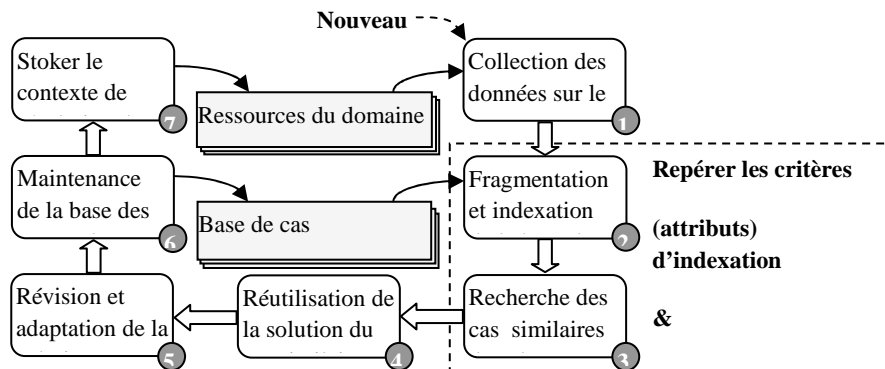


Figure 2 : Les étapes d'intégration des techniques du Datamining et des VLDB dans le processus du RàPC

Conclusion

Le RàPC a apporté un courant d'air frais, et un certain degré d'optimisme dans le secteur de l'intelligence artificielle en général et des systèmes de gestion des connaissances en particulier. Le RàPC, comme une méthodologie de conception des systèmes intelligents, a montré une grande flexibilité d'intégration des solutions techniques issues de deux domaines qui émergent actuellement, le Datamining et les bases de données avancées, lors de l'implémentation des systèmes conçus. Après avoir présenté un état de l'art des travaux réalisés concernant l'intégration des solutions du Datamining dans le processus du RàPC, et les possibilités d'utilisation des méthodes d'optimisation, de visualisation et d'interrogation empruntées du domaine des bases de données avancées dans le cycle du RàPC ; on a proposé dans cet article une démarche pour l'intégration de divers algorithmes du Datamining dans chaque étape de la méthodologie RàPC, renforcée par une optimisation de la recherche des cas similaires, en la limitant dans le fragment pertinent de la base des cas déjà fragmentée.

On finira par mentionner les perspectives très prometteuses du couplage du RàPC avec les solutions du Datamining d'un côté, surtout dans la phase de recherche des

cas similaires ; et des techniques issues des VLDB d'un autre côté, surtout dans la phase de préparation et visualisation des données, vues les diverses caractéristiques communes entre les bases de données et les bases de cas ; on cite par exemple : la représentation multidimensionnelle, l'indexation et les vues matérialisées sur les bases de cas ; la recherche des règles d'associations entre les cas de la base pour analyser l'enchaînement du raisonnement par cas.

Références

- [1] Arshadi, N., Badie, K. A "Compositional Approach to Solution Adaptation in Case-Based Reasoning and its Application to Tutoring Library", Proceedings of 8 th German Workshop on CBR, Lammerbuckel, 2000.
- [2] Bellatreche L., "Utilisation des vues matérialisées, des index et de la fragmentation dans la conception logique et physique d'un entrepôt de données", Thèse doctorale. Université Clermont-Ferrand II, 2000.
- [3] Chang P-C, Liu C-H, Lai R-K. "A fuzzy case-based reasoning model for sales forecasting in print circuit board industries", Expert Systems with Applications 34, 2008, 2049–2058.
- [4] Cortes Robles G., "Management de l'innovation technologique et des connaissances : Synergie entre la théorie TRIZ et le Raisonnement à Partir de Cas. Application en génie des procédés et systèmes industriels", Thèse doctorale. Institut National Polytechnique de Toulouse, 2006.
- [5] Devèze B., Fouquin M., "Case-Based Reasoning", Rapport SCIA, école EPTA, France, 2004.
- [6] Dieng-Kuntz R., Corby O., Gandon F., Giboin A., Golebiowska J., matta N., Ribière M., "Méthodes et outils pour la gestion des connaissances : une approche pluridisciplinaire du Knowledge management", Dunod, 2ème édition, Paris, 2001.
- [7] Duizabo S., Guillaume N., "Les problématiques de gestion des connaissances dans les entreprises", Centre de recherches DMSP (Dauphine Marketing Stratégie Prospectiv), Université Paris Dauphine. France, 1997.
- [8] [Ermine J-L., "Enjeux, démarches et processus de la gestion des connaissances". Support de cours, université de trier, France, 2005.
- [9] Gilleron R., Tommasi M., "Découverte de connaissance à partir de données". Technical report, Grappa, Université Lille 3, France, 2000.
- [10] Kim K-S., Han I., "The cluster-indexing method for case-based reasoning using self-organizing maps and learning vector quantization for bond rating cases". Expert systems with Application 21, 2001, 147-156.

- [11] Leake D-B., Wilson D-C. "When Experience is Wrong : Examining CBR for Changing Tasks and Environments? ". Computer Science Department, Indiana University, U.S.A, 1999
- [12] Liu D-R, Ke C-K., "Knowledge support for problem-solving in a production process: A hybrid of knowledge discovery and case-based reasoning", *Revue : Expert Systems with Applications*, 2006.
- [13] "Data warehouse and data mining", Disponible sur : <http://home.nordnet.fr/~dnakache/valeur>.
- [14] Pan R, Yang Q, Pan S-J., "Mining competent case bases for case-based reasoning". *Artificial Intelligence* 171, 2007, 1039–1068.
- [15] Rakoto H., "Intégration du Retour d'Expérience dans les processus industriels. Application à Alstom Transport", Thèse doctorale. Institut National Polytechnique de Toulouse, 2004.
- [16] Rasovska I., "Contribution à une méthodologie de capitalisation des connaissances basée sur le raisonnement à partir de cas. Application au diagnostic dans une plateforme d'e-maintenance". Thèse doctorale. L'UFR des Sciences et Techniques de l'Université de Franche-Comté, 2006.
- [17] Shiu S., Yeung D., Sun C., Wang X., "Transferring Case Knowledge To Adaptation Knowledge : An Approach for Case-Base Maintenance", *Computational Intelligence*, Volume 17, Number 2, May 2001, pp. 295-314(20).
- [18] Watson I., "Case-based reasoning is a methodology not a technology", *Revue : AI, CBR*, University of Salford, United Kingdom, 1999.
- [19] "Very large database", Disponible sur : http://en.wikipedia.org/wiki/Very_large_database.
- [20] Yang Q., Wu J., "Keep it simple: A case-base maintenance policy based on clustering and information theory", *Proceedings of the Canadian AI Conference*, 2000, p. 102–114.
- [21] Zhuang Z-Y., Churilov L., Burstein F., Sikaris K., "Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners", *Revue Européenne de la recherche opérationnelle*, 2007.