



HAL
open science

Vers une approche générique du raisonnement par cas : application à la gestion énergétique dans le bâtiment

Boulmaiz Fateh, Patrick Reignier, Ploix Stephane

► To cite this version:

Boulmaiz Fateh, Patrick Reignier, Ploix Stephane. Vers une approche générique du raisonnement par cas : application à la gestion énergétique dans le bâtiment. CIGI Qualita MOSIM 2023 × Propulser la performance Interconnectivité et collaboration dans un contexte d'intelligence artificielle, Université de Québec, Jun 2023, Montréal, Canada. hal-04382085

HAL Id: hal-04382085

<https://hal.science/hal-04382085>

Submitted on 9 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CIGI QUALITA MOSIM 2023

Vers une approche générique du raisonnement par cas : application à la gestion énergétique dans le bâtiment

FATEH BOULMAIZ¹, PATRICK REIGNIER¹, STEPHANE PLOIX¹

¹Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, G-SCOP
38000 Grenoble, France
prenom.nom@uni-grenoble-alpes.fr

Résumé – Le raisonnement à base de cas (CBR) est sans doute l'une des approches les plus intuitives de résolution de problèmes en intelligence artificielle. Elle est inhérente à la réutilisation de l'expérience existante qui comprend des solutions aux problèmes ou des mécanismes pour dériver ces solutions. Malheureusement, les approches CBR existants manquent de généralité puisque le processus d'adaptation est guidé généralement par le domaine d'application et repose largement sur les connaissances de l'expert du domaine. De plus, les systèmes CBR existants sont peu performants car ils traitent chaque étape de la méthodologie CBR séparément et indépendamment des autres étapes. Ce travail vise à poser les fondations d'un système générique pour le paradigme CBR en proposant une approche non-dépendante du domaine. Chaque étape du processus de raisonnement est conçue pour répondre non seulement aux exigences spécifiques de l'étape mais aussi pour soutenir les processus des autres étapes. Les résultats expérimentaux de l'application de l'approche proposée dans le cadre d'un système de gestion d'énergie (EMS) accompagnant les occupants de bâtiment vers de meilleures pratiques énergétiques montrent l'efficacité de notre approche.

Abstract – Case-based reasoning (CBR) is arguably one of the most intuitive approaches to problem solving in artificial intelligence. It is intrinsic to the reuse of existing experience that includes solutions to problems or mechanisms to derive those solutions. Unfortunately, existing CBR approaches lack generality, as the adaptation process is typically guided by the application domain and relies heavily on the knowledge of the domain expert. Moreover, existing CBR systems are poorly performing as they handle each step of the CBR method separately and independently from the other steps. This work aims at laying the foundations of a generic system for the CBR paradigm by proposing a domain independent approach. Each step of the reasoning process is designed to meet not only the specific requirements of that step but also to support the processes of the other steps. Experimental results of applying the proposed approach in the context of an energy management system (EMS) guiding building occupants towards better energy practices show the effectiveness of our approach.

Mots clés – raisonnement par cas, similarité, adaptation, algorithme génétique, gestion énergétique dans les bâtiments.

Keywords – case-based reasoning, similarity, adaptation, genetic algorithm, energy management in buildings.

1 INTRODUCTION

Le raisonnement à base de cas est une discipline de l'intelligence artificielle dont les applications sont nombreuses dans des domaines aussi variés que l'ingénierie (Irgens, 2000), la médecine (Zato, 2013), la finance (Luigi Portinale and Giorgio Leonardi and Paolo Artusio, 2016), etc. L'approche CBR est inspirée du raisonnement humain qui utilise les expériences passées de résolution de problèmes pour adresser un nouveau problème.

La méthode CBR a, dès ses débuts, suscité un grand enthousiasme au sein de la communauté de l'intelligence artificielle, dans la mesure où elle était considérée comme un moyen efficace de résoudre de nombreux problèmes posés, jusqu'à présent, par l'approche classique à base de règles et les techniques alternatives telles que les réseaux de neurones ou les modèles régressifs. Plusieurs considérations motivent cet engouement. Premièrement, l'approche CBR permet de surmonter le goulot d'étranglement de l'acquisition de connaissances qui a pesé sur les approches d'apprentissage automatique depuis leur introduction. Deuxièmement, CBR peut faciliter la maintenance du système (par exemple, la

modification des règles de raisonnement) en évitant l'effet domino qui constitue un défi pour certaines approches basées sur des règles. Enfin, contrairement à d'autres approches d'apprentissage automatique, un système CBR peut raisonner en fournissant des solutions acceptables même s'il ne dispose pas d'un ensemble de données suffisamment important et correct pour garantir un apprentissage optimal. En effet, le focus sur les connaissances spécifiques à une situation comme ressource de référence pour le raisonnement facilite considérablement le processus d'acquisition de connaissances.

Malgré les avantages offerts par l'approche CBR, cette dernière a introduit de nouveaux défis qui lui sont propres. Elle a facilité l'acquisition de connaissances à un certain niveau puisque les cas sont faciles à apprendre, mais il ne s'agit que d'une catégorie de connaissances nécessaires pour la CBR. D'autres types de connaissances sont nécessaires pour accomplir le processus de raisonnement, notamment la description du domaine, les connaissances de similarité et les connaissances d'adaptation. Chaque type de connaissance implique un processus d'apprentissage spécifique. Contrairement aux connaissances du domaine et aux connaissances du cas, les connaissances de similarité et d'adaptation disponibles initialement sont difficiles

à modéliser, imprécises ou incomplètes, de plus, elles peuvent évoluer dans le temps. Il est donc important de disposer de mécanismes d'aide à l'acquisition de ces connaissances pour les affiner et les faire évoluer au fur et à mesure de l'utilisation du système. Ceci pose le défi de la gestion de ces bases de connaissances depuis leur conception jusqu'à leur maintenance en passant par le processus d'acquisition. Des travaux antérieurs ont abordé ce problème en considérant chaque type de connaissance de manière autonome. Cependant, ces connaissances sont étroitement liées et ne doivent pas être considérées indépendamment. Par exemple, la connaissance de l'adaptation est souvent utilisée dans l'étape de recherche, afin de réduire l'effort d'adaptation.

Par ailleurs, malgré l'abondante littérature sur le sujet, les stratégies d'adaptation existantes sont fortement dépendantes du domaine d'application, ce qui rend le déploiement massif des systèmes existants difficilement envisageable et les processus de maintenance complexes.

Nous présentons dans cette étude un Framework performant pour l'approche CBR basé sur une méthodologie holistique. Plus précisément, les principales contributions de cette étude sont récapitulées ainsi :

- Proposer un formalisme plus précis que le formalisme traditionnel (problème, solution) pour représenter un cas.
- Développer une approche basée sur un algorithme génétique pour la pondération des variables caractérisant un cas.
- Fournir une méthode pour identifier les cas similaires en définissant des métriques appropriées basées sur le nouveau formalisme de présentation de cas et la métrique statistique F1-score.
- Proposer une nouvelle stratégie d'adaptation visant à développer un processus de réutilisation efficace et aussi indépendant que possible du domaine d'application.

Cet article est organisé comme suit. Après une introduction au contexte de cette étude dans la section 2, la section 3 fournit un aperçu des travaux existants en liens avec la problématique de cet article. Notre méthodologie CBR est expliquée dans la section 4. La section 5 présente une évaluation de l'approche proposée en considérant un cas d'étude avec des données du monde réel. Enfin, la section 6 conclut et propose des pistes de réflexion pour améliorer l'approche proposée.

2 CONTEXTE

Le CBR est une approche de résolution de problèmes qui exploite des expériences antérieures (appelées cas sources) pour résoudre de nouveaux problèmes (cibles). L'ensemble des expériences constitue une base de cas. Typiquement, un cas est constitué d'au moins deux parties : une description de la situation représentant un problème et une solution appliquée pour résoudre cette situation. Le principe sous-jacent à l'approche CBR repose sur l'hypothèse de cohérence qui stipule que *les problèmes similaires ont des solutions similaires*.

Un système CBR est un écosystème basé sur les connaissances combinant plusieurs processus pour mémoriser et exploiter les expériences passées. Pour des raisons de formalisation, les approches d'acquisition de connaissances supposent que les connaissances sont structurées dans des conteneurs de connaissances distincts et que le processus de raisonnement est décomposé en différentes étapes. Cette distinction est utile pour une meilleure

compréhension du mécanisme CBR, bien qu'elle ne représente pas la réalité. En effet, les processus de raisonnement et les conteneurs de connaissances sont fortement dépendants les uns des autres et, en pratique, il arrive souvent que deux ou plusieurs de ces composants soient combinés en un seul processus

2.1 Le processus de raisonnement

Le processus de raisonnement se déroule en cinq étapes.

- **Élaboration** : Les données permettant de formuler un problème sont recueillies et structurées de manière à constituer un nouveau cas. L'environnement du système est sollicité pour recueillir les connaissances requises pour formuler le problème cible à résoudre, ainsi que toutes les connaissances nécessaires pour lancer le processus de raisonnement.
- **Remémoration** : L'objectif est d'identifier dans la base de cas un ou plusieurs problèmes similaires au problème cible, généralement par l'utilisation de connaissances de similarité.
- **Réutilisation** : Lors de l'étape de remémoration, il est rare d'obtenir des cas sources qui correspondent au problème cible, l'étape de réutilisation vise à adapter les solutions des cas sources similaires pour répondre aux exigences du problème cible en s'appuyant généralement sur des connaissances d'adaptation.
- **Révision** : L'objectif de cette étape est d'évaluer la pertinence de la solution proposée pour résoudre le problème cible. Précisément, la révision consiste à fournir au système un retour sur la qualité de la solution et à expliquer les causes en cas d'échec de manière à enrichir les connaissances du système.
- **Mémorisation** : Les compétences du système sont améliorées par la mémorisation du nouveau cas appris. D'autres informations comme les traces du processus de raisonnement peuvent également être mémorisées pour améliorer les connaissances d'adaptation et de similarité.

2.2 Les connaissances

La présentation des différentes phases du cycle de résolution de problème dans CBR a souligné la nécessité de mobiliser différents types de connaissances. L'auteur de (Richter, 1995) propose d'organiser ces connaissances en quatre conteneurs. 1) *les connaissances du vocabulaire* qui définissent les variables modélisant les phénomènes pertinents au domaine d'application ; 2) *les connaissances d'un cas* est une caractéristique du paradigme CBR qui le distingue des autres approches de l'IA en matière de résolution de problèmes. La connaissance de cas est

une connaissance particulière à une situation, obtenue par la résolution de problèmes dans le passé; 3) *les connaissances de similarité* sont utilisées pour sélectionner les cas à réutiliser dans une situation particulière, c.-à-d., estimer l'utilité des cas sources par rapport à la tâche actuelle de résolution de problème; 4) *les connaissances d'adaptation* définissent la manière dont une ou plusieurs solutions identifiées peut être adaptée aux contraintes particulières d'une nouvelle situation à résoudre.

Bien que cette classification présente les différents types de connaissances séparément, il est important de souligner l'étroite interdépendance entre les conteneurs de connaissances. En particulier, l'établissement d'une relation entre la connaissance de similarité et celles d'adaptation est sans aucun doute nécessaire pour une performance optimale du système CBR.

3 ETAT DE L'ART

De nombreux travaux ont été proposés dans la littérature visant à développer des méthodes d'adaptation génériques. Dans une approche telle que celle adoptée dans (Miller & Bergmann, 2015) et (Cordier, Fuchs, & Mille, 2006), l'apprentissage des connaissances d'adaptation consiste à générer des règles d'adaptation en attribuant les différences observées dans les attributs des solutions des cas aux différences dans les attributs des problèmes correspondants. Cependant, la modélisation des corrélations entre les descripteurs du problème et ceux de la solution n'est pas toujours facile et nécessite une connaissance approfondie du domaine d'application. Une approche différente (Hanney & Keane, 1996) a proposé d'apprendre les connaissances d'adaptation sous forme de règles à partir de la généralisation des connaissances des cas. L'inconvénient de cette méthode est qu'elle déplace la maintenance des connaissances d'adaptation vers une base de règles, avec tous les problèmes de maintenance que cela implique.

Une autre direction de recherche propose d'apprendre automatiquement les connaissances d'adaptation en utilisant des techniques d'apprentissage automatique. Les modèles proposés dans (Policastro, Carvalho, & Delbem, 2008) et (Jalali, Vahid; Leake, David, 2015) exploitent un modèle de régression linéaire pour générer des règles d'adaptation. Plus récemment, des travaux tels que (Liao, Liu, & Chao, 2018) et (Ye, Leake, Jalali, & Crandall, 2021) proposent d'apprendre les connaissances d'adaptation en entraînant des réseaux de neurones pour établir une correspondance entre les différences entre les problèmes des cas et les différences entre les solutions correspondantes.

Des études préliminaires sur le CBR (Stahl, 2004) (Kolodner J. L., 1991) ont constaté la difficulté de l'apprentissage des connaissances d'adaptation. Par conséquent, les efforts se sont concentrés sur la fourniture de techniques visant à rendre le processus d'adaptation aussi facile que possible. Une direction de recherche largement adoptée consiste à guider le processus de similarité par l'adaptation. Dans ce cas, les mesures de similarité tiennent compte du critère d'adaptabilité des cas. Les partisans de cette approche soutiennent que ce ne sont pas nécessairement les cas les plus similaires au cas cible qui sont les plus faciles à modifier. Les auteurs de (Smyth, Barry; Keane, Mark T., 1998) proposent d'étendre les mesures de similarité classiques avec des connaissances d'adaptation pour évaluer si

un cas est facilement modifiable pour satisfaire un problème cible. Le concept est d'identifier un critère capable de prévoir l'adaptabilité d'un cas sans avoir à l'adapter complètement (filtrage de cas). Il est clair que le succès de l'utilisation de la méthode de recherche guidée par l'adaptation dépend de la disponibilité des connaissances explicites d'adaptation et de critères d'adaptabilité bien spécifiés. En outre, il existe des domaines où le filtrage des cas est soit irréalisable, soit imprécis, car il n'existe pas de mécanisme fiable pour évaluer l'adaptabilité. La méthode proposée dans (Rousu & Aarts, 1996), suggère de renforcer le processus de remémoration avec une fonction qui estime le coût de l'adaptation liée à chaque cas utilisé dans l'adaptation. Cependant, la définition d'une fonction de coût d'adaptation n'est pas facile et elle est fortement liée au domaine d'application.

Une autre direction de recherche (Fuchs, Lieber, Mille, & Napoli, 2000) exploite les chemins de similarité pour effectuer une adaptation compositionnelle. Un chemin de similarité consiste en une succession linéaire de sous-problèmes liés par des relations. Chaque relation est modélisée par un processus d'adaptation particulier qui permet de passer d'un sous-problème à un autre. Malheureusement, l'apprentissage de ce dernier n'est pas toujours simple et dépend du domaine d'application. La littérature sur l'utilisation de l'optimisation pour dériver la solution du cas cible est riche, par exemple (Azzeh, Nassif, & Banitaan, 2014). Cependant, l'idée d'appliquer l'optimisation à l'adaptation, telle que présentée dans cet article, n'a pas encore été explorée.

Le processus de remémoration est largement exploré dans la littérature en s'appuyant sur des mesures de similarité. En particulier, l'utilisation d'approches de type KNN est populaire dans la recherche de cas sources similaires. Dans ces approches, on choisit un nombre constant de voisins prédéterminé au détriment d'un nombre optimal spécifique à chaque cas cible. Cela a un impact significatif sur la précision du processus de remémoration puisque les cas cibles peuvent ne pas avoir un nombre identique de cas sources similaires. Peu de chercheurs ont tenté de proposer des approches qui adaptent le nombre de cas sources similaires en fonction du contexte du cas cible. Les auteurs dans (Park, Kim, & Chun, 2006) décrivent une méthode permettant de définir le nombre de voisins à partir d'un seuil de distance de similarité en étudiant la distribution des distances entre toutes les paires de cas sources. Le nombre de cas sources similaires respectant la distance de similarité seuil est déterminé par rapport à une probabilité de seuil définie par un expert. Cependant, le fait de s'appuyer sur la distribution des distances entre les cas sources n'est pas en soi représentatif de la similarité entre ces cas. L'étude actuelle s'appuie sur les prémisses de ce travail mais propose une approche plus précise pour définir la distance seuil de similarité.

4 APPRENDRE A MIEUX RAISONNER

Nous présentons une nouvelle méthodologie pour répondre aux constats exposés dans l'introduction. En particulier, nous présentons un nouveau formalisme pour représenter un cas, suivi par la description d'une nouvelle approche pour le processus de remémoration en proposant une nouvelle méthode pour évaluer la similarité et enfin nous développons une approche indépendante du domaine pour le processus d'adaptation.

4.1 Représentation de cas

Souvent, les approches existantes de CBR adoptent la structure de cas proposée initialement dans (Kolodner J., 1993), où les connaissances du cas sont décomposées en une spécification du problème et une description de la solution. La partie problème

décrit les objectifs à atteindre, tandis que la partie solution décrit le processus proposé par le mécanisme de raisonnement pour résolution du problème. Contrairement aux approches existantes et pour les besoins de conception des différentes phases de notre méthodologie, nous définissons une nouvelle structure pour un cas en proposant une classification plus fine des connaissances du domaine d'application, qui sont supposés être décrits dans un langage L_D , comme suit :

- *Variables de contexte* (V_C) modélisant les phénomènes subis par l'environnement du domaine d'application, ce sont des phénomènes sur lesquels on n'a pas de contrôle. Par exemple, les conditions météorologiques dans un EMS.
- *Variables d'actions* (V_A) représentant les phénomènes contrôlables. Par exemple, le point de consigne d'un système de chauffage dans un EMS.
- *Variables d'effets* (V_E) modélisant l'état du système après l'application des actions. Par exemple, la température à l'intérieur de bâtiment après l'ouverture des fenêtres dans un EMS.

Nous supposons que les variables de contexte, d'actions et d'effets sont décrites dans des langages L_C , L_A et L_E respectivement, de sorte que $L_D = L_C \cup L_A \cup L_E$.

Définition 1 : Un cas \mathbb{C} est un tuple de trois éléments $(C, A, E) \in L_C \times L_A \times L_E$ qui suppose l'existence d'une relation $R: L_C \times L_A \rightarrow L_E$, ce qui signifie que E est la conséquence de l'application de A à C .

En considérant la structure revisitée d'un cas introduite par la *Définition 1*, l'hypothèse de principe du paradigme CBR (des problèmes similaires produisent des solutions similaires), est reformulée par l'*Hypothèse 1*.

Hypothèse 1 : Effectuer les mêmes actions dans des contextes similaires génère des effets similaires.

Dans ce qui suit, nous supposons que chaque variable $Q \in \{V_C, V_A, V_E\}$ est représentée par un descripteur $d_Q = (v_Q, R_Q)$, où v_Q est un attribut unique associé à un sous-langage $R_Q \in \{L_C, L_A, L_E\}$ et R_Q peut prendre plusieurs formes. Par exemple une valeur atomique, sous la forme de contraintes, ou plus généralement, sous la forme d'un vecteur $\{x_1, x_2, \dots, x_n\} \subset L_Q$. Dans notre approche, nous adoptons la dernière représentation qui est particulièrement appropriée aux situations dont la théorie de domaine est faible.

4.2 Apprendre à remémorer avec davantage de précision

Le principal objectif de l'étape de remémoration est d'identifier les cas pertinents pour le processus de résolution du problème en cours. Dans la suite, nous introduisons d'abord quelques définitions et notations qui seront employées tout au long de la présentation de notre approche, puis nous décrivons la méthode proposée pour estimer les poids des variables, et nous terminons cette section par une présentation de l'approche de remémoration de cas similaires.

4.2.1 Définitions et notations

Définition 2 : La similarité S est une fonction U qui implique une relation de préférence binaire \succsim définie sur une base de cas non vide CB telle que pour un cas cible \mathbb{C}_T , pour chaque $(\mathbb{C}_1, \mathbb{C}_2) \in CB$, $U(\mathbb{C}_T, \mathbb{C}_1) \geq U(\mathbb{C}_T, \mathbb{C}_2)$ si $\mathbb{C}_1 \succsim \mathbb{C}_2$.

Définition 3 : La distance sensible d_S^Q relative à une variable Q correspond à un seuil défini par un expert du domaine et qui représente la distance maximale au-delà de laquelle deux valeurs de la variable Q ne sont pas considérées comme similaires.

En pratique dans l'évaluation de la similarité, il est rare que les valeurs d'une variable soient identiques, intuitivement, la notion

de distance sensible permet d'identifier les cas les plus similaires possible sans pour autant qu'ils soient identiques. Par exemple, le changement de température de l'environnement est perceptible par l'homme à partir d'un seuil de 1°C. Plus généralement, nous adoptons la notion de distance sensible pour comparer les variables d'effet entre deux cas.

Définition 4 : Deux cas sont similaires du point de vue de leurs effets si la distance maximale entre les valeurs de chaque variable d'effet des deux cas ne dépasse pas la distance sensible correspondante : $\forall (\mathbb{C}_1, \mathbb{C}_2) \in CB, \forall E_i \in L_E, E_i^{\mathbb{C}_1} - E_i^{\mathbb{C}_2} \leq d_S^{E_i}$. Avec - $d_S^{E_i}$ est la distance sensible de la variable d'effet E_i .

Définition 5 : La performance $P_{C_i}(C_i, A_i)$ d'un cas $\mathbb{C}_i(C_i, A_i, E_i)$ modélise la qualité des effets engendrés par l'application d'un plan d'actions à un contexte particulier, c'est-à-dire la (dé)satisfaction $S_{C_i}(E_i)$ de l'utilisateur des effets générés par l'application des actions A_i au contexte C_i .

A noter que par la suite, aucune hypothèse n'est faite quant aux métriques de distance utilisées. Toute métrique de distance peut être utilisée dans la mesure où elle supporte le formalisme introduit dans la section 4.1.

Les constats tirés de la littérature (Keyvanpour & Tavoli, 2012), (Iqbal, 2010) confirment que l'intégration de la pondération des variables dans le calcul de la distance est essentielle vu qu'elle permet d'obtenir des résultats plus précis. Ainsi, nous proposons d'adopter une métrique qui tient compte des poids des variables dans son évaluation de la distance. Par la suite, nous présentons une méthode compétente utilisant un algorithme génétique (AG) basé sur une stratégie de clustering pour estimer les poids des variables.

4.2.2 Pondération des variables

Le principe de notre approche de pondération consiste à déterminer les poids qui permettent de grouper les cas les plus similaires (ayant les distances action-contexte D_{CA} les plus proches) et, par conséquent, ceux dont les distances effet D_E sont les plus proches. Il est fondamental d'associer les variables d'actions et de contexte (à travers la distance D_{CA}) puisque les effets des actions dépendent du contexte dans lequel elles sont appliquées et vice versa. Par exemple, pour un EMS, les conséquences (effets) de l'ensoleillement sur la température d'une pièce en hiver dépendent des actions d'ouverture/fermeture des stores. Les principales étapes du processus de l'AG sont décrites brièvement comme suit :

1. *Création de population initiale de chromosomes* : La population initiale de l'algorithme génétique est générée aléatoirement. Elle est composée d'un certain nombre d'individus qui sont composés de valeurs réelles correspondant aux poids recherchés. La dimension de chaque individu correspond au nombre de variables de contexte et d'actions dans la base de cas.
2. *Groupement des cas similaires* : Appliquer un processus de clustering K-means sur les cas de la base de cas en utilisant la distance D_{CA} pondérée par les valeurs des chromosomes.
3. *Évaluation des performances* : Évaluer la performance de chaque individu conformément à la fonction de fitness qui consiste à calculer la moyenne de la distance D_E entre les cas du même cluster. Le critère d'optimisation est composé de deux termes. Le premier consiste à minimiser la moyenne de la distance D_E tandis que le second consiste à minimiser le nombre de clusters formés.
4. *Génération d'une nouvelle population* : Application des opérateurs d'évolution (sélection, croisement, mutation) pour produire une nouvelle population.

5. Réitération des étapes 2 à 4 jusqu'à la satisfaction du critère de fin. Pour des raisons d'efficacité de calcul, le processus d'optimisation se termine lorsqu'aucune amélioration du critère d'optimisation n'est observée après 10 itérations consécutives ou lorsque le nombre d'itérations atteint 500 itérations.

4.2.3 Recherche de cas similaires

À la différence des méthodes traditionnelles basées sur la distance pour déterminer les cas similaires, qui reposent sur une approche de type KNN, nous introduisons une approche qui adapte automatiquement le nombre de cas sources similaires en fonction d'un seuil de distance de similarité. L'évaluation de la similarité est basée sur la distance D_{CA} pour les mêmes motifs expliqués dans la section 4.2.2. L'objectif est de déterminer les cas sources similaires à partir d'une métrique de similarité, qui considère une distance seuil au-delà de laquelle les cas ne sont pas considérés similaires.

$$\forall \mathbb{C}_i, \mathbb{C}_j \in CB, S(\mathbb{C}_i, \mathbb{C}_j) = \begin{cases} \left(1 - \frac{D_{CA}(\mathbb{C}_i, \mathbb{C}_j)}{\mu}\right), & \text{if } D_{CA}(\mathbb{C}_i, \mathbb{C}_j) \leq \mu \\ 0, & \text{sinon} \end{cases} \quad (1)$$

L'équation (1) définit une mesure de similarité qui satisfait ce critère. Elle donne une signification à μ , il s'agit de la distance D_{CA} maximale qui ne doit pas être dépassée entre deux cas considérés comme similaires. L'apprentissage de la distance seuil μ_* pour le cas cible \mathbb{C}_i se devise en deux étapes :

- *Apprendre un seuil de distance pour chaque cas source* : Il n'est pas possible de formuler une approche commune pour l'apprentissage d'un seuil de distance permettant d'identifier les cas satisfaisant le critère de similarité, cette distance est dépendante du domaine et soumise au contexte de chaque cas. Étant l'Hypothèse 1, nous introduisons la *Propriété 1*, qui définit le seuil de distance $\mu_{\mathbb{C}_i}$ relatif à chaque cas source \mathbb{C}_i . On note $\Pi = \{\mu_{\mathbb{C}_i}\}, \forall \mathbb{C}_i \in CB$.

Cette étape génère, au pire, un nombre de seuils de distance équivalent au nombre de cas sources $|\Pi| = |CB|$

- *Apprendre un seuil de distance pour le cas cible* : Le but est d'estimer un seuil de distance μ_* pour le cas cible \mathbb{C}_* à partir des distances seuils calculées dans l'étape précédente. Trouver le seuil de distance μ_* consiste à rechercher, dans l'ensemble Π , la distance donnant le meilleur compromis entre le rappel (le ratio des cas similaires pertinents extraits par rapport au nombre total de cas similaires pertinents) et la précision (le nombre de cas similaires pertinents extraits par rapport au nombre total de cas extraits). Pour ce faire, la méthode décisionnelle statistique F1-score, est utilisée. Ainsi, le seuil de distance μ_* correspond à la distance $\mu_{\mathbb{C}_i} \in \Pi$ maximisant le F1-score comme exprimé dans l'équation (2).

$$\mu_* = \operatorname{argmax}_{\mu_{\mathbb{C}_i}} (F1 - \operatorname{score}(\mu_{\mathbb{C}_i})) \quad (2)$$

Propriété 1 : Le seuil de distance $\mu_{\mathbb{C}_i}$ pour chaque cas \mathbb{C}_i doit satisfaire la condition : $\forall \mathbb{C}_j \in CB, D_{CA}(\mathbb{C}_i, \mathbb{C}_j) \leq \mu_{\mathbb{C}_i} \Rightarrow \mathbb{C}_i$ et \mathbb{C}_j sont similaires à base de leurs effets (voir Définition 4).

Cependant, il n'est pas possible d'appliquer le seuil contexte-action μ_* de manière directe. En effet, les actions du cas \mathbb{C}_* sont inconnues (c'est précisément ce que nous recherchons) et la seule information disponible pour le nouveau cas \mathbb{C}_* est la donnée de contexte. Par conséquent, les variables de contexte sont les seules données utilisées pour évaluer la similarité du cas \mathbb{C}_* avec ceux de la base de cas en utilisant la distance de similarité de contexte D_C . Dans ce cas, le profil de la fonction de remémoration est le suivant :

$$F_S : C_* \mapsto (C_i, A_i, E_i) \stackrel{\text{def}}{=} \mathbb{C}_i \in CB \quad (3)$$

Le défi qui se pose à ce stade est de définir le seuil $\bar{\mu}_*$ qui représente la distance contexte maximale définissant les voisins similaires $S_{\mathbb{C}_*}$ au cas \mathbb{C}_* , avec $\forall \mathbb{C}_i \in S_{\mathbb{C}_*}, D_C(\mathbb{C}_*, \mathbb{C}_i) \leq \bar{\mu}_*$.

Une façon de surmonter cette difficulté est de déduire le seuil $\bar{\mu}_*$ de la distance contexte à partir du seuil μ_* de la distance D_{CA} en utilisant une fonction de projection ϕ qui transforme l'espace des variables $L_{C^n} \times L_{A^m}$ en un sous-espace des caractéristiques à n dimensions $\phi(L_{C^n} \times L_{A^m}) \in L_C^n$, où n , le nombre de variables de contexte et m , le nombre de variables d'actions. La fonction de projection ϕ a pour objectif de conserver la structure des données entre l'ensemble de données original et les données résultant de la transformation. La structure des données est capturée sous forme de distances entre les cas. Nous ne présentons pas le processus de projection faute de place, le lecteur trouve dans (Espadoto, Martins, Kerren, & Hirata, 2021) une présentation exhaustive sur le sujet.

4.3 Apprendre à mieux réutiliser

Le processus d'adaptation consiste à modifier une ou plusieurs actions sources similaires pour générer les effets cibles satisfaisant les contraintes du contexte cible. Formellement, le processus d'adaptation est un couple (F_S, F_A) , où F_S est une fonction de remémoration entre les cas sources et le contexte du cas cible comme le présente la Fonction (3). La fonction F_A est une fonction de transformation qui génère une solution potentielle A_* en modifiant les solutions des cas sources obtenues grâce à la fonction F_S . Le profil de la fonction F_A est donné par la formule :

$$F_A : (L_C \times L_A \times L_E)^k \times L_C \rightarrow L_A \\ ((C_i, A_i, E_i)_{i \in \{1, k\}}, C_*) \rightarrow A_*, \forall \mathbb{C}_i \stackrel{\text{def}}{=} (C_i, A_i, E_i) \in S_{C_*} \quad (4)$$

Avec $k = |S_{C_*}|$ le nombre de cas sources similaires au cas \mathbb{C}_* .

La formule (4) n'impose aucune contrainte quant au nombre de cas similaires requis pour effectuer l'adaptation, à condition qu'il ne soit pas nul. L'approche que nous proposons dans cet article est une méthode d'adaptation compositionnelle, où les solutions de plusieurs cas sources sont efficacement combinées pour produire une solution cible. En effet, des études antérieures (Chedrawy & Raza Abidi, 2006) ont souligné l'avantage de l'adaptation compositionnelle par rapport à l'adaptation à cas unique qui génère généralement des solutions moins précises. Ceci est dû au fait que souvent, seule une partie du problème source est pertinente pour le problème cible, ce qui rend la tâche d'adaptation difficile voire impossible.

La stratégie d'adaptation proposée ici assume une approche d'adaptation en deux phases :

1. *Apprentissage d'un modèle hybride* : Une des principales raisons de la faiblesse du processus d'apprentissage dans les systèmes d'apprentissage automatique est de favoriser soit un apprentissage exclusivement local, soit un apprentissage purement global. Dans une stratégie entièrement locale, il est généralement accordé une importance excessive aux irrégularités locales des données d'entrée et, en conséquence, le processus d'apprentissage ne parvient pas à saisir la structure des données. Dans une stratégie strictement globale, la complexité du modèle peut s'avérer inadéquate sur certaines parties de l'ensemble des données d'entrée. Il est évident que les deux stratégies d'apprentissage sont complémentaires.

Pour profiter des avantages de chacune de ces stratégies, il serait souhaitable de proposer une approche d'apprentissage hybride qui permette de faire converger la dichotomie entre l'apprentissage global et l'apprentissage local. Nous proposons d'apprendre un modèle à partir des cas similaires (voisins) identifiés dans l'étape précédente en utilisant les

données pondérées. Ce modèle n'est pas strictement local car il capture la structure globale des données en tenant compte de l'influence des poids attribués aux variables.

La littérature propose de nombreuses approches pour l'apprentissage de modèles, notamment les réseaux de neurones artificiels, les arbres de décision, les machines à vecteurs de support, les régressions gaussiennes, les réseaux bayésiens, etc. Néanmoins, il n'existe pas de critères clairs pour le choix d'un modèle. En outre, aucune de ces techniques ne se distingue par son efficacité par rapport aux autres. Le modèle adopté reste le choix du concepteur. L'étude des algorithmes d'apprentissage dépasse le cadre de ce travail, on peut trouver davantage de détails sur les modèles d'apprentissage dans (Shalev-Shwartz & Ben-David, 2014).

Plus formellement, cette étape consiste à apprendre un modèle M_{C_*} du comportement local des effets en fonction du contexte et des actions :

$$M_{C_*}: E_i = g(C_i, A_i), \forall \mathbb{C}_i \stackrel{\text{def}}{=} (C_i, A_i, E_i) \in S_{C_*} \quad (3)$$

Avec g est la fonction de modélisation, S_{C_*} est l'ensemble des cas similaires.

2. *Optimisation du modèle* : La solution à proposer (les actions) consiste à résoudre un problème d'optimisation avec une fonction objectif représentant les performances $P(C_*, A_*)$ des actions proposées (cf. Définition 4). Nous décrivons ci-dessous les fondements mathématiques de cette optimisation :

- La fonction objectif, c'est-à-dire le critère qui détermine la meilleure solution à retenir, consiste à maximiser la performance P_{C_*} du cas \mathbb{C}_* à résoudre.
- Les variables de décision, qui correspondent aux variables à modifier pour influencer la valeur de la fonction objectif, sont constituées de variables d'action.
- Les contraintes, c'est-à-dire les conditions que les variables de décision doivent vérifier pour que la solution soit acceptée, doivent satisfaire le modèle local défini par la formule (3).

Ainsi, la forme générale du problème d'optimisation qui consiste à déterminer les actions A_* à proposer pour résoudre le cas \mathbb{C}_* est la suivante :

$E_* = \operatorname{argmax}_{A_*} P(C_*, A_*)$, soumis au modèle appris : $(M_{C_*} : E_* = g(C_*, A_*))$.

5 ÉVALUATION EXPERIMENTALE

5.1 Cas d'étude

Une étude de cas est réalisée pour évaluer l'efficacité de l'approche proposée. Elle consiste à appliquer le paradigme CBR dans le cadre d'un EMS visant à améliorer l'efficacité énergétique des bâtiments.

Plus précisément, l'objectif de l'EMS est d'améliorer le confort des occupants à un coût énergétique équivalent ou plus faible si possible. La majorité des travaux sur l'application du paradigme CBR pour améliorer la flexibilité énergétique implémentent des approches de type demande-réponse, de gestion de la demande par un décalage de la charge, des stratégies de délestage en période de pointe, etc. Malheureusement, ces approches sont souvent rejetées par les occupants qui sont souvent exclus du processus de gestion, pourtant certaines études ont montré que les actions des occupants ont un impact crucial sur l'efficacité énergétique des bâtiments (Uddin, Wei, & Chi, 2021). Ainsi, guider intelligemment l'occupant dans ses actions (telles que l'ouverture/fermeture des fenêtres, l'ajustement des points de consigne du climatiseur), peut contribuer à obtenir un meilleur

confort dans un bâtiment au même coût énergétique si ce n'est moins.

Sur la base de notre approche, nous avons développé un EMS dont l'objectif est de sensibiliser l'occupant d'un logement de l'impact crucial de ses actions sur l'efficacité énergétique du bâtiment. L'EMS guide l'occupant en lui proposant un plan d'actions optimal pour améliorer son confort (température intérieure, qualité de l'air, etc.) sans coût énergétique supplémentaire.

La fonction de performance utilisée dans le processus d'adaptation consiste à évaluer l'insatisfaction de l'utilisateur vis-à-vis des effets générés suite à l'application des actions proposées par le EMS. L'insatisfaction de l'utilisateur concernant le confort thermique $S_T^h(T)$ et la qualité de l'air $S_C^h(C)$ à l'heure h sont modélisées par les formules (5) et (6). L'insatisfaction globale S^h est définie comme la moyenne de $S_T^h(T)$ et de $S_C^h(C)$.

$$S_T^h(T) = \begin{cases} 0, & T \in [21, 23] \\ \frac{T - 23}{26 - 23}, & T > 23 \\ \frac{21 - T}{21 - 18}, & T < 21 \end{cases} \quad (5)$$

$$S_C^h(T) = \begin{cases} 0, & C < 500 \\ \frac{C - 500}{1500 - 1000}, & T > 23 \end{cases} \quad (6)$$

5.2 Description du jeu de données

L'évaluation expérimentale de notre approche a été réalisée à partir de données réelles collectées dans un bureau de recherche appartenant à l'Université de Grenoble Alpes situé à Grenoble, France. Le bureau est équipé de 18 capteurs qui ont permis l'enregistrement des données pour la période du 1^{er} avril 2015 au 30 octobre 2016. Les données météorologiques sont fournies par un prestataire de service pour la même période.

Toutes les mesures sont calculées sous la forme d'une moyenne horaire, à l'exception des variables *Ouverture des fenêtres* (v_{13}) et *Ouverture de la porte* (v_{14}) où les valeurs correspondent à la fraction de l'heure durant laquelle la fenêtre (respectivement la porte) était ouverte. Par exemple, si la fenêtre a été ouverte pendant 15 minutes au cours de la $k^{\text{ième}}$ heure de la journée, alors $v_{13} = 60/15 = 0.25$.

Les différentes variables qui modélisent les phénomènes considérés dans cette expérimentation sont :

- *Variables d'effets* : température intérieure (v_1), concentration intérieure en CO_2 (v_2).
- *Variables de contexte* : température du couloir (v_3), luminosité (v_4), radiation solaire (v_5), vitesse du vent (v_6), concentration en CO_2 dans le couloir (v_7), puissance électrique (v_8), température du chauffage (v_9), occupation (v_{10}), nébulosité (v_{11}), température extérieure (v_{12}).
- *Variables d'actions* : ouverture de la fenêtre (v_{13}), ouverture de la porte (v_{14}).

5.2.1 Prétraitement

Comme pour toute approche orientée données, les performances et la précision de notre approche dépendent fortement de la qualité des données utilisées dans la phase d'apprentissage. Pour assurer une qualité optimale des données, une étape de prétraitement est nécessaire. Dans le présent cas d'étude, le processus de prétraitement de données se compose de trois étapes :

- *Nettoyage des données* : Le processus de nettoyage des données vise à renseigner les valeurs manquantes et à éliminer le bruit tout en identifiant les valeurs aberrantes et en les corrigeant. Chaque valeur aberrante ou manquante

est remplacée par la valeur moyenne des voisins gauche et droite dans le vecteur de valeurs de la variable.

- *Normalisation des données* : Les données collectées par les capteurs sont de natures physiques différentes et présentent des magnitudes différentes. Il est nécessaire de transformer ces données pour éliminer les effets des valeurs dominantes. Compte tenu de la nature numérique de toutes les variables, nous avons opté pour la méthode MinMax pour échelonner les valeurs de chaque variable entre 0 et 1.
- *Filtrage des données* : Les données sont collectées en continu. Pour des raisons d'efficacité, nous filtrons les jours où il n'y a personne au bureau (par exemple, les jours de congé et les week-ends), puisqu'aucune action n'est enregistrée ces jours-là et qu'il n'y a donc aucun intérêt à les utiliser dans le processus de résolution des problèmes.

5.2.2 Structure de la base de cas

La base de cas *CB* utilisée ci-après consiste à organiser sous forme de cas les données journalières obtenues à partir de la base de données décrite précédemment.

Le vecteur représentant chaque variable est codé sous la forme d'un vecteur à 24 dimensions, dont les éléments sont les valeurs des variables à chaque heure (de 0:00 à 23:00). Comme le bâtiment étudié est un bâtiment universitaire, nous avons restreint notre analyse aux seules heures de bureau. Ainsi, nous présentons à l'occupant du bureau un plan d'actions pour la plage horaire comprise entre 08:00 et 20:00 (les heures de bureau) car pendant les autres heures, il n'y a personne dans le bureau et donc aucune action à proposer.

5.3 Configuration du banc d'essai

Suite à la phase de prétraitement des données, 98 cas ont été retenus pour l'expérimentation. La base de cas est divisée aléatoirement en deux ensembles disjoints, l'un d'entraînement (74 cas) et l'autre de test (24 cas), soit 75% et 25% respectivement. Dans notre cas, l'ensemble d'entraînement contient les cas sources (la base de cas) et l'ensemble de test constitue les cas cibles.

Pour évaluer la précision de notre approche, nous devons reproduire le comportement du bureau après l'application du plan d'actions proposé. La simulation est réalisée à l'aide du modèle physique du bureau proposé dans (Alyafi, Pal, Ploix, & Reignier, 2017).

L'expérience a été réalisée sur un ordinateur portable MacBook Pro 13 équipé d'un processeur Intel^(R) Core^(TM) i7-8559U 2.70 GHz 2.60 GHz, 16 Go de RAM, équipé de Windows 10 pro 64-bit. Les algorithmes sont implémentés en Python 3.9.

5.4 Évaluation empirique

5.4.1 Pondération des variables et seuil de distance

Les variables sont pondérées en utilisant l'algorithme génétique présenté dans la section 4.2.2. Ce dernier a convergé vers les poids optimaux des variables dans la 11^{ème} génération. Les valeurs présentées dans le tableau 1 indiquent que la variable v_3 (température du couloir) présente le poids le plus élevé, ce qui est en accord avec les résultats du modèle physique du bureau, confirmant l'efficacité de notre algorithme pour estimer les poids des variables.

Variable	v_3	v_4	v_5	v_6	v_7	v_8
Poids	0.2113	0.1024	0.0258	0.0418	0.0406	0.0878
Variable	v_9	v_{10}	v_{11}	v_{12}	v_{13}	v_{14}
Poids	0.1078	0.0395	0.0039	0.0294	0.0157	0.0938

Tableau 1 : Poids des variables.

L'approche proposée pour évaluer la similarité a permis de définir un seuil pour la distance contexte $\bar{\mu}_* = 1.2$.

Malheureusement, en raison du manque de données, ce seuil ne permet d'obtenir des cas similaires que pour 16 cas sur les 24 cas de validation.

5.4.2 Efficacité de l'algorithme de pondération

L'efficacité du processus de pondération est évaluée à travers la précision du modèle local. Cette dernière est estimée en utilisant deux indicateurs statistiques : l'erreur quadratique moyenne (RMSE) et le coefficient de détermination (R^2) entre la valeur calculée par le modèle (température estimée et concentration estimée de CO_2) et la valeur réelle issue de la base de cas. En particulier, nous étudions l'impact du nombre de plus proches voisins à considérer sur la précision du modèle local appris. Nous expérimentons l'influence de plusieurs valeurs du nombre de voisins K à partir de la valeur la plus faible ($K = 2$) jusqu'à l'ensemble de la base de cas ($K = 72$) en passant par le cas particulier $K = K(\bar{\mu}_*)$ correspondant à la distance seuil $\bar{\mu}_*$. Les résultats sont résumés dans le Tableau 2 où les valeurs des indicateurs considèrent la moyenne des valeurs des indicateurs correspondants pour les cas de test retenus (16 cas).

En examinant les valeurs RMSE et R^2 rapportées dans le Tableau 2, on peut observer qu'une corrélation peut être établie entre le nombre de K et les valeurs RMSE et R^2 . En effet, on peut vérifier qu'alors qu'il existe une tendance nette à l'augmentation de R^2 lorsqu'on s'éloigne de la valeur $\bar{\mu}_*$, une tendance claire à la diminution de RMSE peut être observée. Cela souligne le fait que la considération de moins de cas plus similaires est plus appropriée pour le processus CBR que la considération d'un grand nombre de cas moins similaires.

K	Température		Concentration CO_2	
	RMSE	R^2	RMSE	R^2
72	0.659	76.21	18.85	74.09
60	0.619	76.84	15.14	76.28
50	0.602	78.02	14.35	81.92
40	0.559	78.51	12.63	83.78
30	0.553	79.02	9.04	86.97
20	0.531	82.17	8.76	88.64
10	0.315	85.42	5.28	90.96
2	0.273	89.67	4.44	91.66
$K(\bar{\mu}_*)$	0.103	91.85	4.02	92.97

Tableau 2 : Précision du modèle

5.4.3 Performance du processus d'adaptation

L'efficacité de l'approche d'adaptation est évaluée en comparant, pour chaque cas de test, les performances (en utilisant les formules (5) et (6)) des actions proposées et les performances des actions déjà enregistrées dans la base de cas. La moyenne des performances de toutes les variables d'effet est utilisée pour estimer la performance globale.

Les Figures (1.a) et (1.b) montrent les performances du système au regard des variables v_1 et v_2 . Pour les deux variables, le système (barres rouges) surpasse les actions enregistrées (barres grises) pour les 16 cas de test. Plus précisément, le système est plus performant de 13 à 24 % pour la variable v_1 et jusqu'à 100 % pour la variable v_2 . Nous observons que pour la variable v_2 , les actions enregistrées pour les cas 0,1,2,3,4,6,9,11,12,13 et 14 sont déjà optimales, le système est aussi bon que ces actions. Pour les cinq autres cas, les résultats du système sont meilleurs. La figure (1.c) illustre la performance globale du système. Les résultats montrent que le système améliore les performances dans 100% des cas. L'amélioration est comprise entre 13% et 31%.

6 CONCLUSION

Nous avons présenté une nouvelle approche de conception de systèmes de raisonnement à base de cas avec des méthodologies plus génériques. En particulier, nous avons proposé un formalisme plus précis de structuration de cas. Ce dernier est

utilisé pour définir une nouvelle méthode d'estimation de similarité qui utilise un algorithme génétique comme mécanisme de pondération des variables. En outre, nous avons introduit une stratégie d'adaptation indépendante du domaine en combinant des méthodes d'apprentissage automatique et d'optimisation. Nous avons illustré l'efficacité de cette approche par un exemple d'application dans le contexte d'un système de

Jalali, V., & Leake, D. (2015). Enhancing case-based regression with automatically-generated ensembles of adaptations. *J. of Intelligent Information Systems*.
 Jalali, V., Leake, D., & Forouzandehmehr, N. (2016). Ensemble of Adaptations for Classification: Learning Adaptation Rules for Categorical Features. *Case-Based Reasoning Research and Development* (pp. 186–202). Springer.

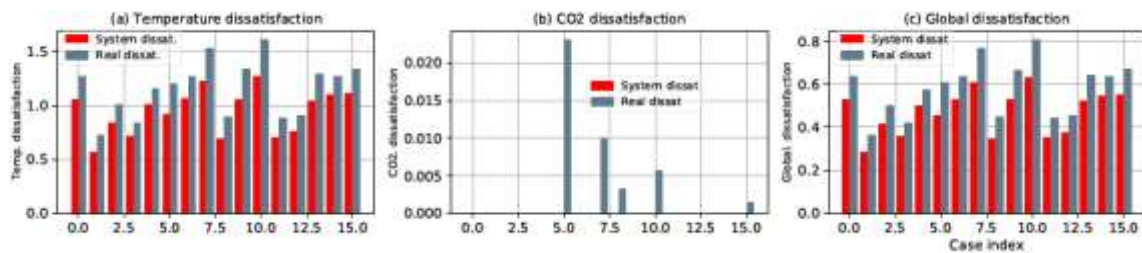


Figure 1 : Performance du processus d'adaptation

gestion de l'énergie. Les travaux futurs comprendront le développement de méthodes plus efficaces pour l'indexation et l'incorporation de nouveaux cas dans la base de cas (l'étape de rétention) en fonction des connaissances acquises lors des étapes de remémoration et d'adaptation. Nous prévoyons également d'étudier la faisabilité d'une approche d'adaptation basée sur l'optimisation en tenant compte de la distance contexte et de la performance de chaque cas voisin. Plus précisément, nous étudierons la performance d'une approche d'adaptation en considérant un opérateur qui non seulement évite les voisins peu performants mais s'en éloigne également.

7 REFERENCES

- Alyafi, A. A., Pal, M., Ploix, S., & Reignier, P. (2017). Differential explanations for energy management in buildings. *2017 Computing Conference*.
- Azzeh, M., Nassif, A. B., & Banitaan, S. (2014). A Better Case Adaptation Method for Case-Based Effort Estimation Using Multi-objective Optimization. *International Conference on Machine Learning and Applications*.
- Bergmann, R., & Schaaf, M. (2003). On the relations between structural case-based reasoning and ontology-based knowledge management. Dans U. Reimer, A. Abecker, S. Staab, & G. Stumme (Éd.), *WM 2003*.
- Chedrawy, Z., & Raza Abidi, S. S. (2006). Case Based Reasoning for Information Personalization: Using a Context-Sensitive Compositional Case Adaptation Approach. *2006 IEEE INES*, (pp. 1-6).
- Cordier, A., Fuchs, B., & Mille, A. (2006). Engineering and Learning of Adaptation Knowledge in Case-Based Reasoning. *Proceedings of the 15th ICMKWN*. Berlin: Springer.
- Espadoto, M., Martins, R. M., Kerren, A., & Hirata, N. S. (2021). Toward a Quantitative Survey of Dimension Reduction Techniques. *IEEE Transactions on Visualization and Computer Graphics*, 27.
- Fuchs, B., Lieber, J., Mille, A., & Napoli, A. (2000, 1). An Algorithm for Adaptation in Case-based Reasoning. *the 14th ECAI-2000*.
- Hanney, K., & Keane, M. T. (1996). Learning adaptation rules from a case-base. Dans *Lecture Notes in Computer Science* (pp. 179–192). Springer Berlin Heidelberg.
- Iqbal, R. (2010, 5). Empirical Learning Aided by Weak Domain Knowledge in the Form of Feature Importance. *Proceedings - 2011 ICMSSP*.
- Irgens, B. L. (2000). Case-based reasoning support for engineering design. (SPIE, Éd.) *Intelligent Systems in Design and Manufacturing III*, 394-402.
- Keyvanpour, M., & Tavoli, R. (2012, 6). Feature Weighting for Improving Document Image Retrieval System Performance. *IJCSI journal*, 9, 125-130.
- Kolodner, J. (1993). *Case-Based Reasoning*. Morgan Kaufmann Publishers.
- Kolodner, J. L. (1991, 4). Improving Human Decision Making through Case-Based Decision Aiding. *AI Mag.*, 12.
- Liao, C.-K., Liu, A., & Chao, Y.-S. (2018). A Machine Learning Approach to Case Adaptation. *2018 IEEE First AIKE*.
- Luigi Portinale and Giorgio Leonardi and Paolo Artusio. (2016). A Smart Financial Advisory System Exploiting Case-Based Reasoning. *CEUR Workshop*.
- Müller, G., & Bergmann, R. (2015). Learning and Applying Adaptation Operators in Process-Oriented Case-Based Reasoning. *ICCBR*.
- Park, y.-j., Kim, B. C., & Chun, S.-H. (2006, 2). New knowledge extraction technique using probability for case-based reasoning: Application to medical diagnosis. *Expert Systems*, 23, 2-20.
- Policastro, C. A., Carvalho, A. C., & Delbem, A. C. (2008, 4). A Hybrid Case Adaptation Approach for Case-Based Reasoning. *Applied Intelligence*, 28, 101–119.
- Richter, M. M. (s.d.). The Knowledge Contained in Similarity Measures. *First Conference, ICCBR-1995, Portugal*.
- Rousu, J., & Aarts, R. J. (1996). Adaptation Cost as a Criterion for Solution Evaluation. *EWCBR-96*. Springer.
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. USA: Cambridge University Press.
- Smyth, B., & Keane, M. T. (1998). Adaptation-guided retrieval: questioning the similarity assumption in reasoning. *Artificial Intelligence*, 102, 249-293.
- Stahl, A. (2004). *Learning of knowledge-intensive similarity measures in case-based reasoning*. Ph.D. dissertation, University of Kaiserslautern.
- Uddin, M. N., Wei, H.-H., & Chi, H. L. (2021). Influence of Occupant Behavior for Building Energy Conservation: A Systematic Review Study of Diverse Modeling and Simulation Approach. *Buildings*, 11.
- Ye, X., Leake, D., Jalali, V., & Crandall, D. J. (2021). Learning Adaptations for Case-Based Classification: A Neural Network Approach. *CBR Research and Development* (pp. 279–293). Springer Publishing.
- Zato, X. B. (2013). *Case-Based Reasoning Applied to Medical Diagnosis and Treatment*. Springer. 2013