

Sujet de thèse en « Configuration Interactive à base de contraintes »

Interactive CONStraint Based Configuration (Iconic)

ISAE SUPAERO / CGI – IMT Mines Albi / IRIT- Université Paul Sabatier

Résumé du projet de thèse

Permettre aux consommateurs ou clients de personnaliser leurs produits (lunettes, chaussures, ordinateurs, voitures, etc.) est l'une des préoccupations actuelles des entreprises, quels que soient leur taille et leur secteur d'activité. Du point de vue des consommateurs, cette personnalisation doit être simple et rapide (quelques clics sur une page Web) tout en permettant d'obtenir le produit correspondant à leurs désirs et leur budget. Du point de vue des entreprises, cette personnalisation repose sur la définition de produits configurables, représentés par des catalogues de composants et leurs relations, et sur la mise en ligne de systèmes de configuration ou configurateurs interactifs.

Les configurateurs raisonnent sur les catalogues de composants (nomenclature générique de la famille de produits) qui sont fortement combinatoires, où les produits et leurs composants sont décrits par un ensemble de contraintes restreignant les valeurs possibles de variables de configurations. L'utilisateur interagit avec le configurateur en sélectionnant des valeurs pour des variables de son choix – le rôle du système étant de ne proposer que des choix garantis compatibles avec les choix courants, les contraintes de l'utilisateur, et de maintenir des indicateurs comme le prix, le délai, etc.

La complexité de ces requêtes (NP ou au dessus) est malheureusement incompatible avec une exécution en ligne. Les méthodes dites « de filtrage » utilisent les contraintes pour effectuer des déductions sur le problème en détectant les valeurs localement incohérentes. Elles garantissent l'interaction avec l'utilisateur mais pas le retrait de toutes les valeurs conduisant à des non-solutions (elles sont donc « incomplètes »). Les méthodes dites « de compilation » proposent de prétraiter le problème original sous une forme qui permet un traitement efficace des requêtes. Cette compilation s'effectue hors ligne, avant la phase de requêtes, ce qui relâche les contraintes sur leur complexité temporelle mais repousse la difficulté dans l'espace : la forme compilée peut avoir une complexité spatiale théorique exponentielle.

Le sujet de thèse proposé a pour but de proposer, concevoir, développer et tester une méthode de raisonnement exploitant conjointement les méthodes de filtrage et celles de compilation. Le développement et les tests de cette nouvelle méthode d'exploitation hybride permettraient de tirer parti du meilleur des deux approches suscitées en atténuant leurs limites identifiées en configuration de produits interactive.

Cette thèse s'effectuera sous une co-direction entre l'ISAE SUPAERO (Pr. Elise Vareilles, DISC), IMT Mines Albi (Paul Gaborit, CGI) et de l'Université Paul Sabatier (Hélène Fargier, DR CNRS, Institut de Recherches en Informatique de Toulouse). Bien que ce projet soit fondamentalement un projet scientifique, il s'inscrit dans un domaine applicatif riche; le doctorant sera amené à se rapprocher d'équipes de recherche et d'équipes de R&D européennes et françaises, par exemple, au niveau national Renault-SIG et Pros-Toulouse, et au niveau international Technical **University of Denmark** et Siemens. De ces collaborations potentielles découlent des perspectives de débouchés applicatives pour le projet et une bonne intégration du doctorant dans le tissu R&D privé.

Motivations

Dans les approches interactives, c'est l'utilisateur et non la machine, qui résout un problème combinatoire d'optimisation de préférences. Les problèmes de configuration de produits en sont des exemples typiques (on configure interactivement une voiture, un ordinateur, etc). En laissant le client explorer l'offre, la configuration en ligne lui permet de maximiser sa satisfaction.

Les problèmes de satisfaction de contraintes ou « CSP » sont très souvent utilisés dans ces applications, tant en recherche que dans l'industrie [Felfernig+2014]. Les modélisations à bases de contraintes permettent de formaliser facilement le produit configurable par un ensemble de variables, chacune associée à un domaine de définition et reliées entre elles par des contraintes qui limitent les combinaisons de valeurs autorisées. Elles permettent ainsi de décrire de manière exhaustive l'espace des solutions, c'est-à-dire l'ensemble des produits possibles. Les CSP présentent plusieurs avantages :

- une grande liberté de modélisation des connaissances (compatibilités entre composants, formules mathématiques d'évaluation, activation, etc.),
- la non-orientation du raisonnement : toute variable présente dans le problème est à la fois variable d'entrée (pouvant être restreinte par l'utilisateur) ou de sortie (résultant d'un calcul, par exemple). Il est donc tout à fait possible de contraindre le prix et d'identifier l'ensemble des produits correspondants,
- une séparation claire entre les modèles de connaissances et leur exploitation (traitement algorithmique),
- la possibilité de combinaisons ou de couplages avec d'autres approches d'exploitation des connaissances, et plus particulièrement avec les raisonnements à partir de cas ou CBR, Data-mining et les ontologies.

Les systèmes de configuration à base de contraintes permettent à un utilisateur de parcourir l'espace de solutions du CSP (l'ensemble des produits potentiels) en lui offrant la possibilité de [Vareilles2015] :

- visualiser la solution en cours de configuration de manière pertinente en ne présentant que les composants, variantes et options effectivement présents dans la solution,
- exprimer ses préférences sur les composants, variantes et options, telles que le choix d'une valeur unique, choix d'ensemble de valeurs, exclusion d'un ensemble de valeurs ou bien l'expression de préférences explicitement entre valeurs,
- d'estimer la solution en cours de configuration selon plusieurs critères parfois antagonistes, tels que le coût, la performance ou le délai de livraison,
- d'exprimer ses contraintes quant aux critères d'évaluation, comme par exemple, limiter le coût / la performance / le délai de livraison de la solution, ce qui limite le choix d'options et de variantes, ou d'optimiser les solutions un de ces critères.

Techniquement, la configuration est un processus itératif qui consiste à retirer de l'espace des solutions (ensemble des produits ou composants possibles), des solutions (produits ou composants) qui ne sont plus cohérentes avec les choix réalisés par l'utilisateur (typiquement, le client) et le modèle générique. Par un processus itératif, l'utilisateur spécifie progressivement ses besoins et converge progressivement vers une/des solution(s) satisfaisant ses besoins.

Deux méthodes concurrentes de traitement des CSP permettent de raisonner sur le modèle générique de manière interactive : les méthodes raisonnant sur le problème décrit par un réseau de variables liées par des contraintes, et celles raisonnant sur une représentation de l'espace des solutions sous forme de graphe.

1/ Méthodes raisonnant directement sur les contraintes :

Les méthodes dites « de filtrage » utilisent localement les contraintes pour détecter des valeurs incohérentes avec les choix courants. L'une des techniques les plus utilisées est la technique de renforcement de la cohérence

d'arc [Montanari1974]. La cohérence d'arc vérifie que toute valeur du domaine d'une variable est compatible avec chaque contrainte prise séparément. Il existe des méthodes de filtrage basées sur la cohérence d'arc spécifiques aux différents types de CSP : les techniques de k-cohérence (arc-cohérence et chemin-cohérence) principalement utilisées dans la cas des CSP discrets ou mixtes [Mackworth1977 ; Bessière+1993 ; Bessière+1994], les techniques de 2 cohérence continue [Faltings1994], de 2B-cohérence [Lhomme1993] ou de Box-cohérence [Benhamou+1994 ; Benhamou1996] pour les CSP à variables continues par exemple.

Toute la difficulté est qu' un filtrage parfait est un problème NP-complet dans le cas général – les algorithmes de filtrage de CSP se limitent donc à des raisonnement locaux, approximatifs, mais polynomiaux. En configuration interactive, l'utilisation de méthodes de filtrage garantit l'interaction avec l'utilisateur mais ne garantit donc pas le retrait de toutes les valeurs conduisant à des non-solutions.

Avantage des méthodes de filtrage :

Raisonnement sur tout type de problèmes (discret, continu ou mixte avec ou variable optionnelle, etc.).

Limite des méthodes de filtrage :

Probable conservation de solutions infaisables et recours au mécanisme de retour-arrière pour restaurer la cohérence.

2/ Méthode raisonnant sur une représentation de l'espace des solutions :

Les méthodes dites de compilation sont basées sur la transformation, par compilation, d'un problème de configuration formalisé comme un CSP en un automate à états finis qui représente ainsi de manière exhaustive l'espace des solutions [Vempaty1992]. Ce type de méthode a l'avantage d'éviter les retours arrière une fois l'automate compilé et résout le problème de qualité de filtrage [Amilhastre+2002]. La compilation d'un problème discret est NP-difficile, mais elle est effectuée hors ligne, et non pas en ligne comme le filtrage, et s'avère en pratique très efficace sur les applications réelles. L'utilisation en ligne de l'automate compilé garantit en temps linéaire que les choix utilisateurs conduisent bien à une solution.

Avantage des méthodes de compilation :

Garantie de l'atteinte d'une solution sans retour-arrière

Limite des méthodes de compilation :

Impossibilité de compiler efficacement certains types de problème (variables continues) et explosion spatiale sur les problèmes à composants optionnels.

Objectifs

L'objectif de la thèse de proposer, concevoir, développer et tester des algorithmes de configuration interactive exploitant conjointement les méthodes de filtrage et celles de compilation. Le développement et les tests de cette nouvelle méthode d'exploitation hybride permettraient de tirer parti du meilleur des deux approches suscitées en atténuant leurs limites identifiées en configuration de produits

Scientifiquement, la première idée est de compiler les sous composants du produit et d'utiliser des algorithmes de filtrage pour propager les choix utilisateurs d'un composant sur l'autre – on évite ainsi l'écueil de l'explosion spatiale due aux composants optionnels. Le second axe de travail portera sur la proposition de structures de données plus adaptées à la présence de variables à domaines continus que les classiques diagrammes de décision et dépassant les niveaux de filtrage par cohérence aux bornes. Plutôt que d'adapter aux variables continues les structures imaginées pour les variables discrètes (c'est possible mais inefficace en discrétisant les domaines), on ré-utiliserait en compilation des outils de représentation issus des travaux sur les domaines continus, par exemple les Q-trees et les R* trees.

Plan de réalisation

Cette thèse s'inscrit dans une collaboration entre l'équipe DISC de l'ISAE SUPAERO, l'axe PSP du CGI d'IMT Mines Albi et l'équipe ADRIA de l'IRIT. Elle se déroulera en trois étapes :

- 1 / Typologie des problèmes de configuration et étude de l'adéquation des deux méthodes de raisonnement interactives concurrentes à leurs caractéristiques,
- 2 / Conception d'une méthodologie d'utilisation conjointe des méthodes de filtrages et de compilation dans un même problème de configuration ; étude de complexité, algorithmique et optimisation,
- 3 / Validation des propositions sur plusieurs cas industriels issus principalement du secteur de l'automobile.

Le doctorant sera directement employé par ISAE SUPAERO pendant 3 ans. Son salaire brut correspond à l'indice 437 - 2048 euros brut /mois. Cette thèse s'inscrit, par ailleurs, dans la chaire Knowledge Compilation d'ANITI. Le temps du doctorant sera donc partagé entre Toulouse et Albi.

Présentation des équipes

Pr. Élise Vareilles est Professeur à l'ISAE SUPAERO depuis janvier 2020. Elle travaille sur le développement d'outils interactifs d'aide à la conception basés sur les connaissances et participe au développement et à l'amélioration du logiciel CoFIADe. Elle a déjà supervisé une dizaine de thèses de doctorat et a été impliquée dans 6 projets de recherches nationaux ou internationaux, toujours avec des partenaires industriels. Elle est le co-auteur de plus de 60 articles dont 17 articles dans des revues Web of Sciences et 40 articles dans des conférences internationales de haut niveau. Depuis janvier 2021, elle anime l'équipe de recherche SD (Systèmes Décisionnels) du DISC à l'ISAE SUPAERO.

Dr. Paul Gaborit est maître assistant à IMT Mines Albi depuis 1999. Il enseigne l'informatique, la programmation, les systèmes d'information, les réseaux, les bases de données. Ses sujets de recherche portent sur l'intelligence artificielle, la planification, l'optimisation, la configuration, la programmation par contraintes. Il a co-supervisé une douzaine de thèses de doctorat. Il est co-auteur de plus de 40 publications et communications. Il a participé à l'organisation de nombreuses conférences scientifiques.

Dr. Hélène Fargier est un chercheur senior de l'équipe ADRIA de l'UMR CNRS IRIT. Ses travaux de recherche portent sur la prise de décision sous incertitudes : au niveau formel (théorie de la décision, préférences et représentation des connaissances), au niveau algorithmique (problème de satisfaction de contraintes, compilation de connaissances) et au niveau applicatif (configuration de produits). Depuis 2005, Mme Hélène Fargier a participé à l'encadrement de 10 thèses et a co-publié plus d'une centaine d'articles dont 35 dans des journaux internationaux et 45 dans des conférences internationales sélectives à comité de programme. Mme Hélène Fargier a participé à une vingtaine de projets nationaux ou internationaux et elle a coordonné le projet ANR BR4CP pour *Business recommendation for configurable products*, impliquant deux laboratoires de recherche et 5 entreprises, dont Renault et Pros. En 2014, Mme Hélène Fargier est devenue une « ECCAI Fellow » et elle est éditrice associée du journal « *Artificial Intelligence* ».

Candidature

Envoyez votre Curriculum Vitae détaillé et vos références à Élise Vareilles, Paul Gaborit et Hélène Fargier en soulignant vos atouts. Vous serez rapidement informé(e) si vous êtes sélectionné(e) pour une présentation orale, qui aura lieu soit à Toulouse, soit par Zoom. Le doctorat débute en novembre 2021 à l'ISAE SUPAERO :

- Elise Vareilles, elise.vareilles@isae-supaero.fr, +33.683.063.854
- Paul Gaborit, paul.gaborit@mines-albi.fr, +33.563.493.076
- Hélène Fargier, fargier@irit.fr, +33.561.558.297