

DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **20 décembre 2024**

Nom de famille et prénom de l'auteur : **Monsieur Théo RABUT**

Titre de la thèse : Optimisation bayésienne avec variables mixtes pour la chimie

Résumé



Chaque réaction chimique doit être optimisée avant son industrialisation. L'objectif de cette optimisation est de trouver, en réalisant des expériences, les paramètres réactionnels (e.g. température, concentration, pression) qui minimisent ou maximisent un objectif (e.g. rendement, sélectivité). La conduction de ces expériences nécessite une grande quantité de connaissances expertes. Malgré cela, il n'est raisonnablement pas possible d'établir une forme analytique des relations entre les paramètres réactionnels et les objectifs. Nous parlons donc de l'optimisation d'une boîte-noire. La réalisation d'expériences peut s'avérer coûteuse en termes matériels, temporels ou simplement économiques. Motivés par les contraintes écologiques et économiques qui s'imposent à l'industrie, les travaux décrits dans ce manuscrit s'inscrivent dans un paradigme particulier, à savoir, l'optimisation d'une boîte-noire avec le minimum d'évaluations possibles. Récemment, l'optimisation Bayésienne s'est établie comme la méthode de référence pour l'optimisation de boîte-noire chère à évaluer. Toutefois, il existe de nombreuses variantes au problème d'optimisation de réaction classique. Les travaux exposés dans ce manuscrit traitent un cas particulier où, les paramètres réactionnels sont de types différents. Nous nous intéressons aux variables continues comme le choix de la température de réaction et aux variables catégorielles comme le choix de la base ou du solvant. Ces dernières variables, discrètes et non-ordonnées, sont par leur nature non-numérique difficiles à traiter. Ainsi, la première problématique de cette thèse est la proposition d'une méthode d'optimisation adaptée aux variables mixtes pour la chimie. Nous avons rencontré la nécessité de créer des simulateurs de réactions pour pouvoir évaluer la qualité de nos propositions. Nous traitons donc, dans une première partie, la méthodologie et la qualité de ces simulateurs construits à partir de jeux de données représentant des réactions et des formulations chimiques. Dans un deuxième temps, nous exposerons notre proposition pour l'optimisation de réactions chimiques avec variables continues et catégorielles. Notre proposition est composée d'un

processus gaussien construit avec une fonction de covariance tirée de la littérature et d'une comparaison de deux méthodes pour l'optimisation de la fonction d'acquisition. Par la suite, nous proposons une analyse fine des résultats d'optimisation de nos simulateurs. En effet, malgré la nature probabiliste de l'optimisation, nous considérons que, dans des cas rares, la méthode d'optimisation suggère des expériences redondantes, voir inutiles. Nous exposons donc une suite de visualisation sur les différentes composantes pour déterminer la cause de ces suggestions. La dernière partie de cette thèse expose une proposition pour l'amélioration de notre méthode d'optimisation en se concentrant sur la partie modélisation qui, dans sa forme actuelle, ne différencie pas les variables catégorielles. Or, nous verrons que ces variables peuvent avoir des impacts différents sur la fonction objectif. Ces travaux s'inspirent d'un mécanisme connu sous le nom de Automatic Relevance Determination dans la communauté de l'apprentissage statistique. Nous appelons donc notre modification d'une fonction de covariance populaire, Categorical Automatic Relevance Determination ou CATARD. À travers ce manuscrit, nous exposons un gain de performances pour l'optimisation de fonctions sous-jacentes aux réactions et formulations chimiques avec des paramètres continus et catégoriels. De plus, un regard particulier est tourné vers d'autres applications de notre méthode d'optimisation dans la mesure où il existe de nombreux problèmes analogues à notre problème d'optimisation de réaction chimique à variables mixtes.

Mots-clés : Optimisation, Apprentissage automatique, Processus chimique, Boîte-noire,