



HAL
open science

Design de plan expérimental optimal : vers une approche basée sur la logique floue

Olivier Rousselle, Jean-Philippe Poli, Nadia Ben Abdallah

► To cite this version:

Olivier Rousselle, Jean-Philippe Poli, Nadia Ben Abdallah. Design de plan expérimental optimal : vers une approche basée sur la logique floue. LFA 2023 - Rencontres francophones sur la logique floue et ses applications, Nov 2023, Bourges, France. Cepaduès, Rencontres francophones sur la logique floue et ses applications, pp.97-100, 2023, Rencontres Francophones sur la Logique Floue et ses Applications 2023. cea-04410406

HAL Id: cea-04410406

<https://cea.hal.science/cea-04410406>

Submitted on 22 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.

Design de plan expérimental optimal : vers une approche basée sur la logique floue

Optimal design of experiments: towards a fuzzy logic approach

O. Rousselle¹

J.P. Poli¹

N. Ben Abdallah¹

¹ Université Paris-Saclay, CEA, List, F91120 Palaiseau, France

olivier.rousselle@cea.fr, jean-philippe.poli@cea.fr, nadia.benabdallah@cea.fr

Résumé :

Nous présentons dans cet article les fondements d'une méthode basée sur la logique floue pour le design de plan expérimental, sous contraintes et avec peu de données. L'objectif est de fournir aux expérimentateurs un algorithme interprétable / explicable, leur permettant d'échantillonner de manière optimale. Nous détaillons les différentes étapes de notre algorithme (génération du système de règles, inférence floue, ...), les résultats obtenus, ainsi que la comparaison avec l'optimisation bayésienne.

Mots-clés :

Règles floues, inférence, optimisation, interprétabilité.

Abstract:

We present in this article the foundation of a method using fuzzy logic for experimental plan design under constraints and with few data. The goal is to provide experimenters with an interpretable / explicable algorithm, allowing them to sample optimally. We detail the different steps of our algorithm (generation of the rule system, fuzzy inference,...), the results obtained, and the comparison with Bayesian optimization.

Keywords:

Fuzzy rules, inference, optimization, interpretability.

1 Introduction

La science des matériaux est le croisement de différentes disciplines dans le but de produire des matériaux performants pour un usage donné. Ces matériaux sont généralement produits à partir d'un mélange de composants initiaux soumis à un certain procédé de fabrication (par ex. fabrication additive, dépôt physique ou chimique de vapeur, etc.) ; depuis quelques années, l'intelligence artificielle accélère l'innovation dans ce domaine [1, 2]. Classiquement, l'approche essais-échecs est utilisée dans ce domaine afin de collecter des

données expérimentales qui sont souvent coûteuses et longues à produire.

Dans ce contexte, l'objectif de notre travail est de développer une méthode basée sur la logique floue appliquée aux problèmes de design de plan expérimental. On définit notre problème comme tester différents jeux de paramètres (par ex. la composition d'un matériau) afin de maximiser une quantité donnée (par ex. la robustesse de ce matériau). En particulier, nous nous intéressons à trouver une méthode automatique pour échantillonner les paramètres expérimentaux de manière optimale.

Différentes méthodes d'échantillonnage sont aujourd'hui utilisées dans les domaines expérimentaux, comme l'échantillonnage aléatoire, factoriel, la méthode de surface de réponse, l'optimisation bayésienne [3], l'algorithme de couverture optimale [4], etc... Les réseaux de neurones peuvent également être utilisés pour la découverte de nouveaux matériaux [5], mais ont l'inconvénient d'être une boîte noire. Notre motivation est de fournir un outil pour aider les expérimentateurs à déterminer quels sont les prochains jeux de paramètres à tester alors qu'ils disposent de peu de données, et pour comprendre pourquoi ce choix de paramètres a été fait par l'algorithme (interprétabilité). En effet, une telle approche permettrait de réduire le gaspillage des matières premières et de converger plus vite vers un matériau innovant. L'explicabilité / interprétabilité vise à rendre le fonctionnement et les résultats du modèle

davantage intelligibles et transparents pour les humains, afin de renforcer la confiance en la prise de décision. Dans notre cas d'étude, les variables d'entrées sont dans un premier temps des données compositionnelles x_i de mélange (avec la contrainte $\sum x_i = 1$), et la variable de sortie y est la propriété à optimiser.

Par exemple, en science des matériaux, il s'agit d'avoir en entrée les quantités de produits bruts à mélanger et en sortie une performance du matériau à atteindre, comme sa résistance aux corrosions, etc. Dans nos récents travaux, nous nous intéressons en particulier à utiliser de telles approches pour construire les batteries de nouvelle génération. Dans ce cas, la composition pourrait être la quantité de silicium, argent, bismuth, et la performance à maximiser serait les capacités de charge / décharge.

Cet article présente les premiers pas de nos travaux dans ce domaine et le cheminement que nous proposons pour atteindre notre objectif.

L'article est décomposé comme suit. La section 2 présente nos motivations pour l'élaboration d'un modèle basé sur la logique floue. La section 3 détaille les différentes étapes de l'algorithme, et la section 4 présente les premiers résultats obtenus sur un jeu de données simulé. Nous terminons par des conclusions et perspectives.

2 Motivations

Afin de satisfaire les besoins des expérimentateurs, notre approche doit répondre à différents critères :

- Echantillonnage adaptatif [6], i.e. processus séquentiel qui décide de l'emplacement de l'échantillon suivant en équilibrant des critères d'exploration et d'exploitation ;
- Combinaison apprentissage (données expérimentales) et connaissances expertes ;
- Fiabilité et robustesse : des petits changements dans les points initiaux ne doivent pas entraîner de grands

changements dans les résultats et prédictions ;

- Explicabilité et interprétabilité : le fonctionnement et les résultats du modèle doivent être intelligibles pour les humains/expérimentateurs ;
- Passage à l'échelle : notre approche doit pouvoir fonctionner pour des problèmes de mélange à grande dimension ;
- Efficace en termes de temps de calcul.

Dans la section suivante, nous établissons les principes de notre approche.

3 Présentation de l'approche

Afin de satisfaire ces critères, nous avons développé une approche dont nous détaillons ici les différentes étapes :

- 1) Apprentissage du vocabulaire flou à partir d'un jeu de données limité. Il s'agit de déterminer automatiquement les partitions floues pour les différentes entrées utilisées dans les règles. En effet, à terme, nous proposerons cela lorsque l'expertise n'est pas disponible pour réaliser ces partitions manuellement. La méthode de partition choisie repose sur un clustering flou, où un point appartient à chaque cluster avec un poids compris entre 0 et 1. Les fonctions d'appartenance obtenues ont une forme trapézoïdale. En l'absence d'un nombre suffisant de données, cette étape est sous-optimale et devra être repensée.

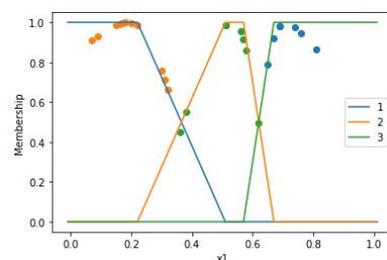


Figure 1 – Exemple de partition forte pour la variable d'entrée x_1 . Les points représentent les degrés d'appartenance obtenus par le clustering flou.

- 2) Génération d'un système de règles floues grâce à l'algorithme de Wang et Mendel [7,

8] : pour chaque point du jeu de données déjà échantillonné (x_1, y) , on détermine pour chaque variable la partition avec le degré d'appartenance maximum; le regroupement de ces partitions floues (une par variable) génère une règle. Un système de règles est alors construit, chaque règle floue étant de la forme : « si x_1 est faible et x_2 est important et x_3 est faible, alors $y = \sum_i \alpha_i x_i + b$ », avec (α_i, b) les coefficients propres à chaque règle et déterminés par régression multi-linéaire à partir des points déjà échantillonnés. Des règles issues de connaissances « expertes » peuvent également être ajoutées à ce système. Ce système de règles est utilisé pour alimenter un système d'inférence flou de type Takagi-Sugeno, qui permet de prédire la valeur de sortie y pour chaque valeur des variables compositionnelles. L'incertitude sur les prédictions est déterminée à partir du degré d'activation de la règle (une prédiction associée à une règle sera d'auto.

- 3) Détermination du prochain point à échantillonner, qui est soit le point optimal qui maximise la sortie prédite (exploitation) ou bien celui associé à l'incertitude la plus élevée (exploration). La valeur de sortie y est évaluée pour ce point, et ce nouveau point est ajouté au jeu de données.

Ce processus peut être répété sur plusieurs itérations (qui correspondent au nombre de répétitions de l'expérience). Les différentes étapes de l'algorithme sont illustrées sur la figure 2.

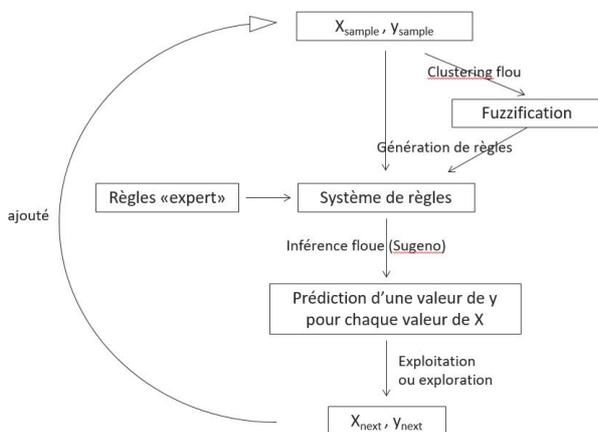


Figure 2 – Illustration des étapes de l'algorithme à base de règles

L'algorithme a été implémenté sous Python 3, avec l'utilisation du package *scikit-fuzzy* pour le clustering flou et l'inférence floue.

4 Résultats préliminaires

L'algorithme a été testé sur un jeu de données simulé dont on connaît la fonction objectif et dont la surface de réponse présente plusieurs extrema locaux (et un extremum global). Un exemple est présenté ci-dessous, où la fonction objectif est un produit de fonctions sinus :

$$f_r(X) = \left| \prod_{i=1}^N \sin(i r \pi X_i) \right|$$

où le paramètre r détermine le nombre d'extremums locaux, et où la valeur de sortie est comprise entre 0 et 1. La surface de réponse de cette fonction est illustrée sur la figure 3.

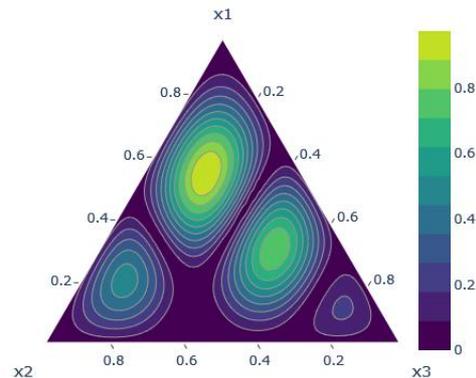


Figure 3 – Surface de réponse de la fonction f_r utilisée pour le test de l'algorithme

Dans cet exemple, cinq points initiaux sont générés aléatoirement dans les zones à faible valeur de sortie (zones en bleu sur la surface de réponse), et 20 itérations sont réalisées.

Les résultats obtenus sont comparés avec l'optimisation bayésienne. L'objectif de notre algorithme est d'atteindre le point optimal en un faible nombre d'itérations, tout en apportant de l'explicabilité.

L'algorithme actuel fournit les résultats résumés dans la figure 4, où sont représentées les meilleures valeurs de sortie échantillonnées en fonction du nombre d'expériences réalisées.

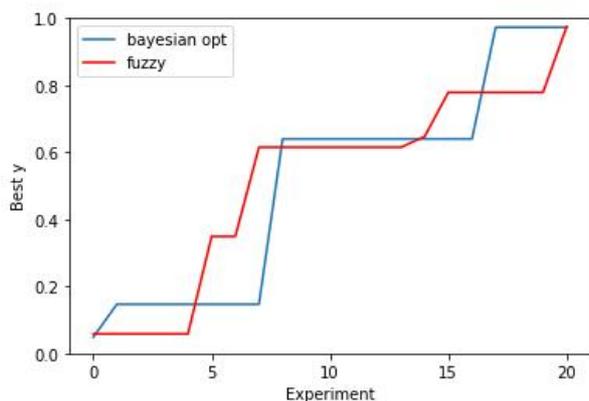


Figure 4 – Performance de l'algorithme d'inférence flou et de l'optimisation bayésienne

L'objectif est d'obtenir à terme un modèle plus performant que l'optimisation bayésienne, plus interprétable, et efficace en terme de temps de calcul. En effet, l'approche de Wang et Mendel que nous avons choisie pour l'apprentissage de règles de régression produit un nombre de règles trop grand, nuisant ainsi à l'interprétabilité, et des conclusions peu représentatives. Toutefois, il permettait de démarrer les travaux par sa simplicité de mise en œuvre.

Cette première version de l'algorithme nous encourage à poursuivre les travaux dans cette direction.

5 Conclusion et perspectives

Les premiers résultats sont prometteurs. Dans notre exemple d'étude, l'algorithme atteint une valeur plus importante que l'optimisation bayésienne. Cette approche a aussi le mérite d'être interprétable, intuitive, et peut être une véritable aide aux expérimentateurs.

Les pistes d'amélioration incluent l'amélioration de la prédiction de la valeur de sortie. A minima, nous pouvons prendre des formes de conclusion de Sugeno plus complexes, en fonction du nombre de données disponibles. Nous envisageons de poursuivre l'étude de l'incertitude associée, puisqu'elle joue un rôle important non seulement dans le choix de la prochaine expérimentation mais

aussi dans le cadre de l'interaction avec l'utilisateur. A ce même titre, nous nous intéresserons à la génération d'une explication. Il faudra également tenir compte des paramètres de procédés (comme la température et la pression) puisque nous les négligeons pour le moment. Enfin, pour que tout cela soit utilisable convenablement, nous regarderons également la vitesse de calcul de l'algorithme. Il serait préférable également de ne pas réapprendre la base de règles à chaque étape, mais de la modifier. Cependant, l'apprentissage continu est un véritable challenge en soi.

Remerciements :

Ce travail est financé par le plan transverse de compétences « Matériaux et Procédés » du CEA. Nous remercions nos collègues F. Schuster, C. Adda, F. Le Cras et G. Yildirim pour leur temps, le partage de leurs connaissances et leurs données expérimentales.

Références

- [1] B. Cao et al., *How To Optimize Materials and Devices via Design of Experiments and Machine Learning: Demonstration Using Organic Photovoltaics*, ACS Nano, 12, 2018
- [2] F. Ren et al., *Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments*, Sci. Adv. 4, 2018
- [3] S. Greenhill et al., *Bayesian Optimization for Adaptive Experimental Design: A Review*, IEEE, 2019
- [4] David S Bem et al, *Combinatorial experimental design using the optimal-coverage algorithm*, 2003
- [5] Y. Liu et al., *Materials discovery and design using machine learning*, Journal of Materiomics, 3, 2017
- [6] D. Xue et al., *Accelerated search for materials with targeted properties by adaptive design*, Nature Communications, 2016
- [7] L.-X. Wang and J. M. Mendel, *Generating fuzzy rules by learning from examples*, Systems, Man and Cybernetics, IEEE Transactions, 22., 1992
- [8] L.-C. Dutu, G. Mauris, P. Bolon, *A Fast and Accurate Rule-Base Generation Method for Mamdani Fuzzy Systems*, IEEE Transactions on Fuzzy Systems, 2018.