Optimisation des interactions homme-machine par l'approche *Bayesian Information Gain*

Culture Sciences de l'Ingénieur

Hugo MIQUEL, Olivier RIOUL et Julien GORI

Édité le 15/12/2025

école normale supérieure paris—saclay

Cette ressource est issue des travaux d'Hugo MIQUEL réalisés dans le cadre de son stage de Master 2 au LTCI (Télécom Paris, Institut Polytechnique de Paris), sous la direction d'Olivier RIOUL (professeur au LTCI) et de Julien GORI (chercheur CNRS à l'ISIR, Sorbonne Université, Inserm).

Cette ressource présente la méthode BIG (*Bayesian Information Gain*), une approche mathématiquement fondée pour minimiser le nombre d'interactions entre l'humain et la machine tout en maximisant leur efficacité informationnelle. S'appuyant sur les principes du BED (*Bayesian Experimental Design*), BIG formalise le problème de l'interaction comme un processus séquentiel d'acquisition d'information optimale sur l'objectif de l'utilisateur. Cette optimisation est particulièrement bénéfique pour les personnes en situation de handicap, pour lesquelles chaque interaction représente un coût significatif. L'article expose les fondements théoriques du BED et leur adaptation au contexte interactif. Des extensions de la méthode sont également présentées pour assurer sa robustesse face aux erreurs de l'utilisateur, garantissant ainsi une expérience interactive fiable même dans des contextes d'utilisation réels. Cette approche ouvre des perspectives prometteuses pour la conception d'interfaces adaptatives minimisant l'effort de l'utilisateur tout en maintenant l'efficacité de la communication homme-machine.

1 - Introduction

La minimisation du nombre d'interactions entre l'humain et la machine représente un enjeu fondamental dans la conception d'interfaces utilisateur efficaces. Cette optimisation est particulièrement cruciale pour les personnes en situation de handicap moteur, pour lesquelles chaque interaction physique avec un dispositif numérique peut représenter un effort considérable. Qu'il s'agisse de personnes atteintes de paralysie, de troubles musculosquelettiques ou simplement de limitations temporaires de mobilité, la réduction du coût d'interaction peut améliorer l'expérience utilisateur et rendre accessibles des technologies autrement difficiles à utiliser.

Imaginons par exemple une personne tétraplégique utilisant un système de contrôle oculaire pour naviguer dans une interface. Chaque sélection représente un effort cognitif et physique important. Dans ce contexte, réduire le nombre d'interactions nécessaires pour atteindre un objectif ne constitue pas simplement une amélioration de confort, mais peut déterminer si la tâche devient réalisable ou non pour l'utilisateur.

Cette problématique s'étend également aux contextes à forte charge cognitive, comme les environnements professionnels complexes où la fatigue décisionnelle peut conduire à des erreurs critiques. Comment alors concevoir des systèmes qui maximisent l'efficacité de chaque interaction tout en minimisant leur nombre?

La réponse se trouve dans le cadre méthodologique BIG (*Bayesian Information Gain*), qui s'appuie sur les principes de la conception expérimentale bayésienne ou BED (*Bayesian Experimental Design*) [3]. Cet article présente les fondements de cette approche et explique comment elle permet d'optimiser les interactions homme-machine [4, 5, 6, 8].

2 - Bayesian Experimental Design : fondements théoriques

La conception expérimentale bayésienne (*Bayesian Experimental Design* ou BED) représente un cadre mathématique rigoureux permettant d'optimiser la conception d'expériences dans des contextes d'incertitude. Développée initialement par Lindley [3] et approfondie par de nombreux chercheurs depuis [1, 9], cette approche se distingue fondamentalement des méthodes fréquentistes classiques par son intégration explicite des connaissances préalables et son traitement probabiliste cohérent des incertitudes.

2.1 - Fondements conceptuels

Contrairement à l'approche fréquentiste qui cherche à optimiser des propriétés moyennes sur de nombreuses répétitions hypothétiques, l'approche bayésienne se concentre sur l'optimisation d'une expérience spécifique, tenant compte des connaissances déjà disponibles.

Le BED repose sur trois composantes fondamentales:

- Une formalisation des connaissances préalables sous forme de distribution de probabilité
- Une définition claire de l'objectif de l'expérience en termes d'utilité espérée
- Un processus d'optimisation du design expérimental pour maximiser cette utilité

2.2 - Formalisation mathématique

Considérons un espace de paramètres Θ dont les éléments $\theta \in \Theta$ représentent les quantités d'intérêt que nous cherchons à estimer ou à mieux comprendre. Dans le cadre bayésien, notre incertitude sur θ est représentée par une distribution de probabilité $p(\theta)$, appelée distribution *a priori*.

Exemple médical: Imaginons un médecin face à un patient présentant certains symptômes. Le paramètre θ pourrait représenter la présence (ou l'absence) d'une maladie particulière. Initialement, le médecin estime une probabilité $p(\theta = \text{malade}) = 0.3$ basée sur la prévalence de la maladie et les symptômes observés. Cette distribution a priori capture l'incertitude initiale du médecin.

Le design expérimental (ou plan d'expérience), noté ξ, représente l'ensemble des choix et paramètres contrôlables qui définissent comment l'expérience sera menée. Il englobe tous les aspects de la conception expérimentale:

- Les conditions dans lesquelles les observations seront faites
- Les moments auxquels les mesures seront prises
- Les stimuli présentés et leur ordre
- Les paramètres d'instruments de mesure
- La sélection des sujets ou unités expérimentales

Dans notre contexte médical, le design expérimental ξ correspond au choix des tests diagnostiques à effectuer. Le médecin peut opter pour:

- ξ_1: un test sanguin standard (peu coûteux mais modérément informatif)
- ξ_2: une imagerie médicale spécialisée (coûteuse mais très informative)
- ξ_3: une combinaison séquentielle de tests en fonction des résultats préliminaires

Chaque design implique différents coûts, délais et valeurs informatives potentielles.

Ce design ξ peut être simple (par exemple, un point unique dans l'espace des paramètres) ou complexe (une séquence de décisions adaptatives). Le choix optimal de ξ dépendra de l'objectif spécifique de l'expérience et des contraintes pratiques.

Le modèle probabiliste de l'expérience est caractérisé par $p(y|\theta, \xi)$, la distribution de probabilité des observations y conditionnellement au paramètre θ et au design ξ. Cette distribution capture la relation entre les quantités d'intérêt et les observations possibles, y compris les erreurs de mesure et autres sources d'incertitude.

Considérons le test sanguin (ξ_1) avec ses caractéristiques connues:

- Sensibilité: $p(y = \text{positif}|\theta = \text{malade}, \xi_1) = 0.9$ (90% des malades sont détectés)
- Spécificité: $p(y = \text{négatif}|\theta = \text{sain}, \xi_1) = 0.8$ (80% des personnes saines ont un résultat négatif)

Ces probabilités conditionnelles définissent $p(y|\theta, \xi_1)$ et intègrent les incertitudes liées à l'imperfection du test.

2.3 - Mise à jour bayésienne et distribution prédictive

Lorsque des observations y sont obtenues sous un design ξ, la connaissance sur θ est mise à jour selon la règle de Bayes [7]:

$$p(\theta|y, \xi) = \frac{p(y|\theta, \xi)p(\theta)}{p(y|\xi)} \tag{1}$$

où $p(\theta|y, \xi)$ est la distribution a posteriori, intégrant à la fois l'information préalable et celle fournie par l'expérience. Le dénominateur $p(y|\xi)$ est la distribution prédictive marginale:

$$p(y|\xi) = \int_{\Theta} p(y|\theta, \xi)p(\theta)d\theta. \tag{2}$$

Cette distribution prédictive joue un rôle crucial dans le BED car elle représente la distribution de probabilité des observations avant leur réalisation effective, tenant compte de toutes les incertitudes sur θ.

Supposons que le médecin choisisse le test sanguin (ξ_1) et obtienne un résultat positif ($y = \text{positif}$). La mise à jour bayésienne donne:

$$\begin{aligned} p(\theta = \text{malade}|y = \text{positif}, \xi_1) &= \frac{p(y = \text{positif}|\theta = \text{malade}, \xi_1) \cdot p(\theta = \text{malade})}{p(y = \text{positif}|\xi_1)} && (3) \\ &= \frac{0.9 \cdot 0.3}{p(y = \text{positif}|\xi_1)}. && (4) \end{aligned}$$

Le dénominateur est calculé comme:

$$\begin{aligned} p(y|\xi_1) &= p(y = |\theta = \text{malade}, \xi_1)p(\theta = \text{malade}) + p(y = |\theta = \text{sain}, \xi_1)p(\theta = \text{sain}) && (5) \\ &= 0.9 \cdot 0.3 + 0.2 \cdot 0.7 = 0.41. && (6) \end{aligned}$$

Ainsi:

$$p(\theta = \text{malade}|y = \text{positif}, \xi_1) = \frac{0.27}{0.41} \approx 0.66. \tag{7}$$

La probabilité que le patient soit malade a augmenté de 30% à 66% après observation du test positif.

2.4 - Critères d'utilité et optimisation du design

L'objectif du BED est de choisir le design ξ^* qui maximise une utilité espérée $U(\xi)$. Mais comment quantifier la valeur d'une expérience avant même de l'avoir réalisée? C'est précisément le rôle du critère d'utilité, qui formalise mathématiquement ce que l'on considère comme une « bonne expérience ». Un critère d'utilité permet de comparer différents designs expérimentaux et de sélectionner celui qui apporte le plus d'information ou qui répond le mieux aux objectifs fixés.

Formellement, le design optimal est celui qui maximise l'utilité espérée:

$$\xi^* = \arg\max_{\xi \in \Xi} U(\xi) \tag{8}$$

où Ξ représente l'ensemble des designs possibles. L'utilité espérée est calculée en moyennant sur toutes les observations possibles:

$$U(\xi) = \int_{\mathcal{Y}} p(y|\xi) U(y, \xi) dy \tag{9}$$

où $U(y, \xi)$ est une fonction d'utilité qui évalue le bénéfice d'obtenir l'observation y avec le design ξ. Cette moyenne pondérée par les probabilités $p(y|\xi)$ garantit que l'on optimise non pas pour un résultat particulier, mais pour tous les résultats possibles de l'expérience.

Comment le médecin peut-il choisir entre les designs ξ_1 (test sanguin) et ξ_2 (imagerie) avant d'observer les résultats?
Supposons que nous utilisions comme critère d'utilité la réduction d'entropie (gain d'information). L'entropie initiale de θ est:

$$H(\theta) = -[0.3 \log(0.3) + 0.7 \log(0.7)] \approx 0.88 \text{ bits.} \tag{10}$$

Pour le test sanguin (ξ_1), les entropies conditionnelles possibles sont:

- Si $y = \text{positif}$ (avec probabilité 0.41): $H(\theta|y = \text{positif}, \xi_1) \approx 0.92$ bits;
- Si $y = \text{négatif}$ (avec probabilité 0.59): $H(\theta|y = \text{négatif}, \xi_1) \approx 0.29$ bits.

L'entropie conditionnelle moyenne est donc:

$$H(\theta|Y, \xi_1) = 0.41 \cdot 0.92 + 0.59 \cdot 0.29 \approx 0.55 \text{ bits.} \tag{11}$$

Le gain d'information espéré est:

$$I(\theta; Y|\xi_1) = H(\theta) - H(\theta|Y, \xi_1) = 0.88 - 0.55 \approx 0.33 \text{ bits.} \tag{12}$$

Un calcul similaire pour l'imagerie (ξ_2), avec sa sensibilité et spécificité plus élevées, pourrait donner:

$$I(\theta; Y|\xi_2) \approx 0.66 \text{ bits.} \tag{13}$$

Dans cet exemple, l'imagerie (ξ_2) fournit plus d'information attendue et serait donc préférée selon ce critère d'utilité, malgré son coût plus élevé. Le médecin pourrait également incorporer le coût financier dans sa fonction d'utilité pour une décision plus équilibrée.

Plusieurs critères d'utilité ont été proposés dans la littérature, chacun reflétant différents objectifs expérimentaux. Nous présentons ici les trois principaux critères utilisés en conception expérimentale bayésienne:

Gain d'information de Shannon: Ce critère, mesure la réduction d'entropie entre la distribution *a priori* et *a posteriori* [2]. L'entropie de Shannon quantifie le degré d'incertitude d'une distribution de probabilité: plus l'entropie est élevée, plus l'incertitude est grande. Le gain d'information mesure donc à quel point une expérience réduit cette incertitude:

$$U(y,\xi) = \int_{\Theta} \log(p(\theta|y,\xi))p(\theta|y,\xi)d\theta - \int_{\Theta} \log(p(\theta))p(\theta)d\theta. \tag{14}$$

Cette mesure est équivalente à l'information mutuelle entre θ et y, qui quantifie la quantité d'information que les observations y apportent sur le paramètre θ:

$$U(\xi) = I(\theta; y) = \int_{\Theta}\int_{\mathcal{Y}} p(\theta, y|\xi) \log \frac{p(\theta, y|\xi)}{p(\theta)p(y|\xi)} dy d\theta. \tag{15}$$

Ce critère est particulièrement adapté lorsque l'objectif est de maximiser la connaissance globale sur tous les paramètres, sans privilégier une région particulière de l'espace des paramètres.

Réduction de variance: Pour les problèmes d'estimation, la réduction de la variance est un critère souvent utilisé. La variance d'une distribution mesure sa dispersion autour de sa moyenne: une variance élevée indique une grande incertitude sur la valeur du paramètre. Ce critère cherche donc à minimiser la dispersion de la distribution a posteriori:

$$U(\xi) = \int_{\mathcal{Y}} p(y|\xi) \left[\mathrm{Var}[\theta] - \mathrm{Var}[\theta|y,\xi]\right] dy. \tag{16}$$

Ce critère est particulièrement utile lorsque l'objectif principal est d'obtenir une estimation précise d'un paramètre numérique.

Divergence de Kullback-Leibler: La divergence de Kullback-Leibler, mesure la "distance" (au sens probabiliste) entre la distribution a posteriori et la distribution a priori. Plus précisément, elle quantifie combien d'information est perdue lorsqu'on approxime la distribution a posteriori par la distribution a priori:

$$U(\xi) = \int_{\mathcal{Y}} p(y|\xi) \int_{\Theta} p(\theta|y,\xi) \log \frac{p(\theta|y,\xi)}{p(\theta)} d\theta dy. \tag{17}$$

Cette mesure est asymétrique et peut être interprétée comme la quantité d'information apportée par l'expérience pour transformer notre connaissance a priori en connaissance a posteriori. Elle est mathématiquement équivalente au gain d'information de Shannon et est donc souvent utilisée de manière interchangeable dans le contexte du BED. Ce critère est particulièrement pertinent lorsqu'on souhaite mesurer l'impact informatif d'une expérience sur nos croyances préalables.

3 - Du BED au BIG: application à l'interaction homme-machine

Le framework Bayesian Information Gain (BIG)[4, 5, 6] adapte les principes du BED au contexte spécifique de l'interaction homme-machine. L'idée fondamentale est de considérer chaque interaction comme une « expérience » visant à réduire l'incertitude sur l'objectif de l'utilisateur.

3.1 - Variables aléatoires et notations

Trois variables aléatoires fondamentales structurent le framework BIG:

- Θ: la cible visée par l'utilisateur (inconnue du système);
- X: le feedback ou la question posée par le système;
- Y: la réponse fournie par l'utilisateur.

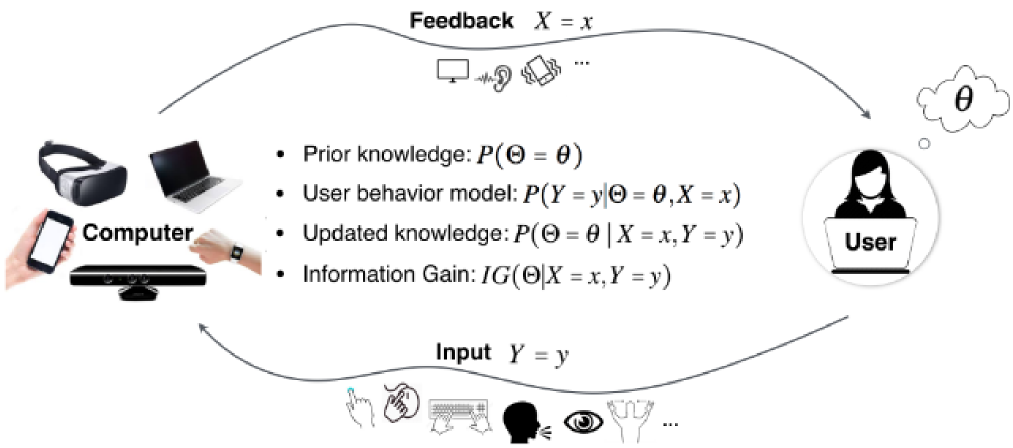

Figure 1: Représentation schématique du framework BIG illustrant les trois variables clés: Θ (cible), X (feedback) et Y (réponse)

3.2 - Optimisation des interactions

À chaque étape, le système doit choisir le feedback x qui maximisera l'information obtenue sur Θ après avoir observé la réponse Y. Cela se traduit par le choix de:

$$x^* = \arg\max_x I(\Theta; Y|X = x) \tag{18}$$

où $I(\Theta; Y|X = x)$ représente l'information mutuelle entre la cible Θ et la réponse Y, étant donné le feedback x.

Cette information mutuelle peut être exprimée comme:

$$I(\Theta; Y|X = x) = H(Y|X = x) - H(Y|\Theta, X = x) \tag{19}$$

où H représente l'entropie de Shannon. Intuitivement, cette équation mesure la réduction d'incertitude sur la réponse Y lorsqu'on connaît la cible Θ, pour un feedback x donné.

3.3 - Gain d'information et mise à jour bayésienne

Le gain d'information quantifie la réduction d'incertitude concernant la cible après une interaction:

$$IG(\Theta|X = x, Y = y) = H(\Theta) - H(\Theta|X = x, Y = y) \tag{20}$$

où $H(\Theta)$ représente l'entropie de Shannon, mesure de l'incertitude sur Θ.

Après avoir observé la réponse y de l'utilisateur, le système met à jour sa croyance sur la cible en appliquant la règle de Bayes:

$$p(\theta|x, y) = \frac{p(y|x, \theta) \cdot p(\theta)}{p(y|x)} \tag{21}$$

Cette distribution a posteriori devient le nouvel a priori pour l'interaction suivante, créant ainsi un cycle d'apprentissage continu.

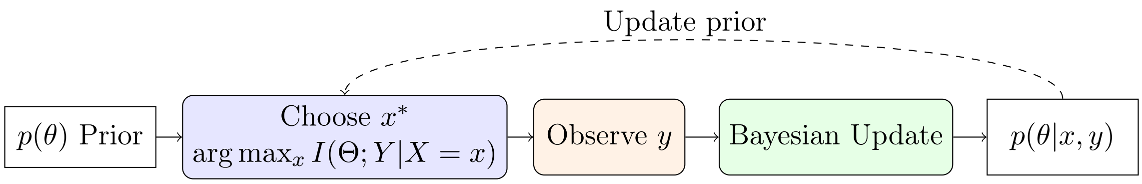

Figure 2: Cycle de mise à jour bayésienne: inférence de θ avec retour direct

4 - Extensions pour la robustesse aux erreurs utilisateurs

Le framework BIG, tel que présenté jusqu'ici, suppose implicitement que l'utilisateur répond toujours correctement aux questions posées par le système. Cette hypothèse, bien que simplificatrice pour la présentation théorique, est rarement satisfaite dans les contextes d'interaction réels. En effet, les utilisateurs peuvent commettre des erreurs pour diverses raisons: fatigue, distraction, mauvaise compréhension de la question, ou contraintes liées à un handicap. Dans ce contexte, le framework BIG est particulièrement vulnérable à ces erreurs, car une seule réponse incorrecte peut éliminer définitivement la véritable cible des hypothèses considérées.

Pour remédier à cette limitation, des extensions ont été développées pour rendre le framework robuste aux erreurs [8]. Un modèle à taux d'erreur fixe intègre un paramètre ϵ_0 représentant la probabilité qu'un utilisateur commette une erreur, permettant ainsi au système de maintenir une probabilité non nulle pour la cible correcte même après une réponse erronée. Une extension plus sophistiquée, le modèle à taux d'erreur adaptatif, traite ce taux d'erreur comme une variable aléatoire à inférer conjointement avec la cible de l'utilisateur. Cette approche permet au système de s'adapter dynamiquement au comportement réel de l'utilisateur, optimisant ainsi l'équilibre entre la précision d'identification et le nombre de questions nécessaires. Ces extensions forment une hiérarchie mathématiquement cohérente, où chaque modèle généralise le précédent, offrant ainsi un cadre théorique unifié pour l'interaction homme-machine robuste aux erreurs.

5 - Conclusion

La méthode BIG (*Bayesian Information Gain*) représente une avancée significative dans l'optimisation des interactions homme-machine. En s'appuyant sur les fondements théoriques solides de la conception expérimentale bayésienne, il permet de réduire le nombre d'interactions tout en maximisant l'efficacité de chaque échange.

Cette approche ouvre des perspectives prometteuses pour l'accessibilité numérique, en rendant les technologies plus accessibles aux personnes en situation de handicap et en améliorant l'expérience utilisateur dans des contextes à forte charge cognitive.

La puissance de BIG réside dans sa capacité à formaliser mathématiquement ce qui intuitivement constitue une « bonne question » qui, quelle que soit la réponse, apportera le maximum d'information pour identifier l'objectif de l'utilisateur. Cette formalisation permet de concevoir des systèmes interactifs fondamentalement plus efficaces et inclusifs.

Références :

[1] Chaloner, K., & Verdinelli, I. (1995). *Bayesian experimental design: A review.* Statistical Science, 10(3), 273-304.

[2] Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory.* John Wiley & Sons.

[3] Lindley, D. V. (1956). *On a measure of information provided by an experiment.* Annals of Mathematical Statistics, 27(4), 986-1005.

[4] Liu, W., d'Oliveira, R. L., Beaudouin-Lafon, M., & Rioul, O. (2017). *BIGnav: Bayesian information gain for guiding multiscale navigation.* Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 5869-5880.

[5] Liu, W., Rioul, O., Mcgrenere, J., Mackay, W. E., & Beaudouin-Lafon, M. (2018). *BIGFile: Bayesian information gain for fast file retrieval.* Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1-13.

[6] Liu, W., Rioul, O., & Beaudouin-Lafon, M. (2021). *Bayesian information gain to design interaction.* Chapter 3, in *Bayesian Methods for Interaction and Design*, Cambridge University Press.

[7] MacKay, D. J. (2003). *Information theory, inference and learning algorithms.* Cambridge University Press.

[8] Miquel H., Gori J., & Rioul O. (2025). *Bayesian experimental design with mutual information and learned errors for human-computer interaction.* European Mathematical Psychology Group Meeting (EMPG 2025), Padua, Italy.

[9] Rainforth, T., Foster A., Ivanova D. R., & Bickford Smith F. (2024). *Modern Bayesian Experimental Design.* Statistical Science, 39(1), 100-114.