Explications contrefactuelles pour les forêts aléatoires prudentes

Haifei Zhang, Benjamin Quost, Marie-Hélène Masson

Explications contrefactuelles pour les forêts aléatoires prudentes

Haifei Zhang[1] Benjamin Quost[1] Marie-Hélène Masson[1,2]

[1] UMR CNRS 7253 Heudiasyc, Université de Technologie de Compiègne, 60200, France
[2] Université de Picardie Jules Verne, IUT de l'Oise, Beauvais, 60000, France

haifei.zhang@hds.utc.fr benjamin.quost@hds.utc.fr mmasson@hds.utc.fr

Résumé :

Les forêts aléatoires prudentes sont conçues pour prendre des décisions indéterminées lorsque l'incertitude est trop forte. Puisque l'indétermination a un coût, il semble cependant souhaitable de comprendre pourquoi une décision précise n'a pu être prise pour une instance particulière. Nous proposons pour cela d'utiliser la notion d'explication contrefactuelle. Nous cherchons quelles modifications minimales peuvent être apportées à une instance pour que la décision indéterminée devienne déterminée. Dans cet article, nous proposons un algorithme efficace pour générer des exemples contrefactuels pour des forêts aléatoires prudentes. Nous évaluons l'efficacité de notre stratégie sur différents ensembles de données et nous l'illustrons sur deux études de cas simples portant sur des données tabulaires et des images.

Mots-clés :

Classification prudente, XAI, exemples contrefactuels.

Abstract:

Cautious random forests are designed to make indeterminate decisions when the uncertainty is too high. Since indeterminacy has a cost, it seems desirable to highlight why a precise decision could not be made for an instance, or which minimal modifications can be made to the instance so that the decision becomes a single class. In this paper, we propose to use the notion of counterfactual and propose an efficient algorithm to generate determinate counterfactual examples for cautious random forests. We evaluate the efficiency of our strategy on different datasets and we illustrate its utility on two simple case studies involving both tabular and image data.

Keywords:

Cautious classification, XAI, counterfactuals.

1 Introduction

Les modèles d'apprentissage automatique atteignent actuellement des performances remarquables dans de nombreux domaines tels que le diagnostic médical, les systèmes de recommandation, la reconnaissance d'images et de la parole. Les sorties de ces modèles sont traditionnellement précises : dans un problème de classification, elles consistent en une classe unique à laquelle l'instance évaluée est affectée. Cependant, lorsque les données d'apprentissage sont peu nombreuses, ou lorsque les erreurs ont un coût très important, les classifieurs prudents, qui fournissent des ensembles de classes plausibles plutôt que des classes uniques, constituent une alternative. Les forêts aléatoires prudentes (FAP) [10] sont l'un de ces classifieurs. Une FAP combine la stratégie classique des forêts aléatoires (FA), le modèle de Dirichlet imprécis (IDM) [9] et la théorie des fonctions de croyance [6]. La différence majeure avec une FA classique est qu'une décision indéterminée peut être prise, en présence à la fois d'incertitude épistémique (lorsque les sorties de l'arbre sont basées sur peu d'exemples d'apprentissage) et d'incertitude aléatoire (le conflit entre ces sorties est important), ce qui se produit généralement près des frontières de décision.

Faire des prédictions imprécises a cependant un coût, puisque l'indétermination doit être levée par une analyse plus approfondie. Par conséquent, il semble crucial de comprendre ce qui a conduit à une décision indéterminée, et ce qui pourrait être fait pour la transformer en une décision déterminée. Ces questions relèvent du thème émergent de l'intelligence artificielle explicable (XAI) [5]. Dans cet article, nous abordons le problème en utilisant des explications contrefactuelles [8] : de telles explications identifient les modifications minimales à apporter à une instance originale x pour la transformer en une instance modifiée x', de sorte que $f(x')$ corresponde à une prédiction souhaitée $y' = f(x') \neq f(x)$.

Notre approche est inspirée de celle proposée par Blanchart [1], spécifiquement développée pour les ensembles d'arbres. Nos contributions consistent en l'amélioration de l'efficacité de la procédure, et en l'exploitation des explications contrefactuelles pour expliquer les sorties indéterminées des FAP. Leur intérêt pour expliquer l'indétermination est illustré par des résultats expérimentaux, en particulier via deux études de cas.

L'article est structuré comme suit. Dans le paragraphe 2, nous rappelons les connaissances générales sur les forêts aléatoires prudentes et les explications contrefactuelles. Leur application à l'explication de l'indétermination est discutée dans le paragraphe 3. Le paragraphe 4 détaille les expériences et discute les résultats. Une brève conclusion est présentée dans le paragraphe 5.

2 Contexte

2.1 Forêts aléatoires prudentes

Les forêts aléatoires prudentes ont été proposées comme une alternative aux forêts aléatoires précises, afin de prendre des décisions à partir de données pauvres — dans ce travail présentes en nombre limité. Dans un problème de classification binaire, pour chaque instance de test x, chaque arbre t de la forêt fournit des éléments d'information sur sa classe réelle $Y \in \{1,0\}$, sous la forme de limites inférieure et supérieure $\underline{p}_1^t(x)$ et $\overline{p}_1^t(x)$ sur la probabilité *a posteriori* $Pr(Y=1|x)$. Ces bornes sont obtenues à l'aide du modèle de Dirichlet imprécis et reflètent l'incertitude de l'estimation due au manque de données d'entraînement. Ces intervalles peuvent être agrégés à l'aide de la théorie des fonctions de croyance, en calculant deux quantités, une croyance et une plausibilité, définies de la manière suivante :

$$\begin{aligned} bel_1(x) &= bel(Pr(Y=1|x) \in [0.5,1]) \\ &= \sum_{[\underline{p}_1^t(x),\overline{p}_1^t(x)] \subseteq [0.5,1]} m_t \\ &= \sum_{t=1}^{T} m_t \mathbb{1}(\underline{p}_1^t(x) \geq 0.5), \end{aligned} \tag{1}$$

et

$$\begin{aligned} pl_1(x) &= pl(Pr(Y=1|x) \in]0.5,1]) \\ &= \sum_{[\underline{p}_1^t(x),\overline{p}_1^t(x)] \cap]0.5,1] \neq \emptyset} m_t \\ &= \sum_{t=1}^{T} m_t \mathbb{1}(\overline{p}_1^t(x) > 0.5), \end{aligned} \tag{2}$$

où m_t $(t = 1,\ldots,T)$ est la masse de l'intervalle $[\underline{p}_1^t(x),\overline{p}_1^t(x)]$ dans le processus d'agrégation et $\mathbb{1}(\cdot)$ est la fonction indicatrice. Cette croyance et cette plausibilité peuvent ensuite être utilisées dans un processus de prise de décision prudente tel que la dominance d'intervalles [10] :

$$\hat{y} = \begin{cases} 1, & \text{si } bel_1(x) \geq 0.5, \\ 0, & \text{si } pl_1(x) < 0.5; \\ \{0,1\}, & \text{sinon.} \end{cases} \tag{3}$$

Comme on peut le voir dans l exemple représenté sur la Figure 1, l'imprécision se produit principalement autour des frontières de décision, où les feuilles de l'arbre ont tendance à contenir peu d'instances et où les sorties de l'arbre sont souvent en conflit les unes avec les autres.

2.2 Explications contrefactuelles

L'intelligence artificielle explicable est un domaine émergent de l'intelligence artificielle qui vise à aider les humains à comprendre les résultats des algorithmes d'apprentissage automatique. Les exemples contrefactuels (CF) sont des explications locales basées sur des exemples, qui peuvent être considérés comme des modifications minimales d'une instance originale x conduisant à des décisions différentes. De tels exemples peuvent être soit

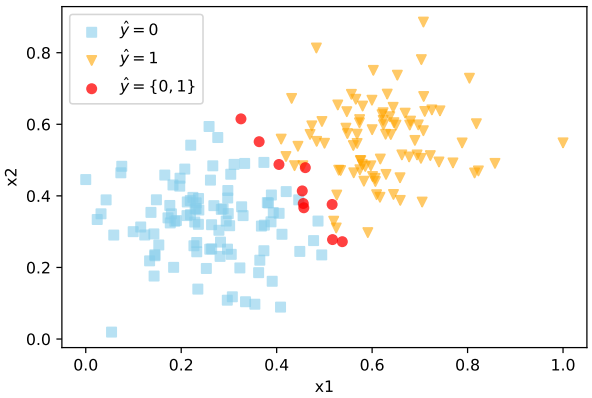

FIGURE 1 – Exemple de prédictions d'un forêt aléatoire prudente.

recherchés dans l'ensemble d'apprentissage, soit synthétisés artificiellement. Étant donné un classifieur f, une requête $x \in \mathscr{X}$, et une étiquette de prédiction souhaitée $y' \in \mathscr{Y}$, le problème est donc de déterminer efficacement un exemple contrefactuel x' en résolvant

$$x' = \arg\min_{z \in \mathscr{X}} \text{dist}(x, z) \quad \text{t.q. } f(z) = y', \qquad (4)$$

où dist est une mesure de distance appropriée (par exemple, euclidienne) entre les instances.

De nombreuses méthodes ont été proposées pour résoudre le problème (4) de manière exacte ou approchée, comme la sélection de l'exemple le plus similaire dans l'ensemble d'apprentissage, la recherche d'un exemple virtuel en optimisant une fonction de coût (pour les modèles différentiables), ou encore des méthodes heuristiques qui peuvent être fondées sur l'approximation du modèle (par exemple par un arbre de décision) de manière à simplifier la recherche [4].

Outre la complexité inhérente aux algorithmes de génération de contrefactuels, d'autres challenges rendent difficile la recherche de contrefactuels : protection de certains attributs ou caractéristiques immuables (comme le sexe, l'ethnicité, etc.), restriction du nombre de caractéristiques modifiées (sparsité) et génération de contrefactuels plausibles (ou réalistes).

3 Expliquer l'imprécision en utilisant des exemples contrefactuels

Ce travail propose d'appliquer les explications contrefactuelles à la classification binaire prudente : étant donné une instance x avec une prédiction indéterminée $f(x) = \{0, 1\}$, nous voulons identifier les deux modifications minimales x^1 et x^0 telles que $f(x^1) = \{1\}$ et $f(x^0) = \{0\}$. Ces deux exemples synthétiques permettront non seulement de mettre en évidence les caractéristiques qui doivent être modifiées pour éliminer l'indétermination, mais aussi de déterminer quelle est l'ampleur de la modification nécessaire pour parvenir à une décision précise.

3.1 Extraction d'exemples contrefactuels pour les forêts aléatoires

Une forêt aléatoire sépare l'espace d'entrée $\mathscr{X}$ en régions de décision, chacune d'entre elles étant elle-même l'intersection de T régions de décision fournies par les T arbres de la forêt. À titre illustratif, la Figure 2 montre le partitionnement de l'espace des caractéristiques par une forêt composée de deux arbres, où les feuilles sont représentées par des boîtes multidimensionnelles, chaque dimension étant découpée en un ensemble d'intervalles.

Pour trouver l'exemple contrefactuel optimal avec la classe désirée y' pour une instance x, il est nécessaire d'examiner toutes les régions de décision de la forêt : la complexité de cette exploration, qui peut se représenter sous forme d'un arbre de recherche, est exponentielle. Cette recherche exhaustive est donc irréalisable pour des données en grande dimension ou des forêts avec des arbres très profonds.

Comme le montre la Figure 3, chaque niveau de l'arbre de recherche correspond à une partition d'une dimension et chacun de ses nœuds est un intervalle, de sorte que le parcours du nœud racine aux feuilles de l'arbre correspond à une région de décision représentée par une boîte multidimensionnelle. Évidemment, il n'est pas

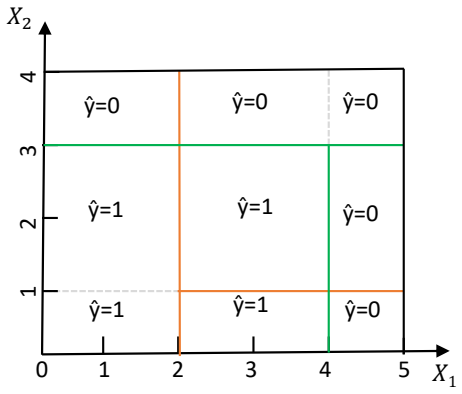

FIGURE 2 – Forêt aléatoire basée sur des données bidimensionnelles et constituée de deux arbres de décision.

nécessaire de développer complètement cet arbre de recherche, mais seulement d'explorer la zone autour de la requête. Blanchart a proposé une stratégie de branch-and-bound à cette fin [1]. Cet arbre de recherche est construit en commençant par la région de décision contenant x, puis en l'étendant à ses voisins en remontant dans l'arbre, dimension par une dimension. Une fois qu'un exemple contrefactuel est trouvé, la distance de la région de décision à la requête fournit une limite supérieure à la distance de la requête à l'exemple contrefactuel, notée d_{sup} (initialisée à l'infini). Sur cette base, les régions de décision R telles que $d(x,R) > d_{sup}$ peuvent être ignorées, où $d(x,R)$ est la distance euclidienne de la requête au point le plus proche dans la région R.

3.2 Filtrage des régions et initialisation contrefactuelle

Étant donnée la complexité de la détermination d'un exemple contrefactuel x' pour une requête x associée à une décision indéterminée $f(x) = \{1,0\}$, nous implémentons deux améliorations pour accélérer la procédure. Ces étapes préliminaires permettent de réduire drastiquement la complexité de la recherche, comme nous le montrerons dans le paragraphe 4.

1. Suivant une suggestion faite dans [1], en présence de caractéristiques protégées (immuables), nous filtrons les régions qui ne correspondent pas aux mêmes valeurs protégées que dans l'exemple x.
2. Nous proposons une approche alternative pour initialiser la recherche d'un exemple contrefactuel, c'est-à-dire pour calculer le premier exemple contrefactuel sur la base duquel le seuil de distance initial d_{sup} sera déterminé (pour remplacer la valeur initiale infinie).

L'intégration de l'étape de filtrage mentionnée ci-dessus dans toute procédure de branch-and-bound est simple. L'étape d'initialisation est critique car elle détermine le seuil d_{sup} et donc le nombre de régions à explorer. Par exemple, dans la Figure 3, si nous savons initialement que d_{sup} est inférieur à 1, alors il n'est pas nécessaire d'explorer les régions $L_2 \cap L_5$ et $L_3 \cap L_6$, puisque leurs distances par rapport à x sont supérieures à 1. Ainsi, le paramètre d_{sup} reçoit une valeur initiale basse, plus nous pouvons filtrer les régions qu'il n'est pas nécessaire d'explorer, et cet avantage est plus important dans les dimensions élevées.

L'approche Minimum Observable (MO), qui sélectionne l'instance la plus proche x' avec la classe désirée y' dans l'ensemble d'apprentissage, est couramment utilisée à cette fin. Cependant, dans les régions de l'espace des caractéristiques pauvres en données d'apprentissage, la distance entre la requête et

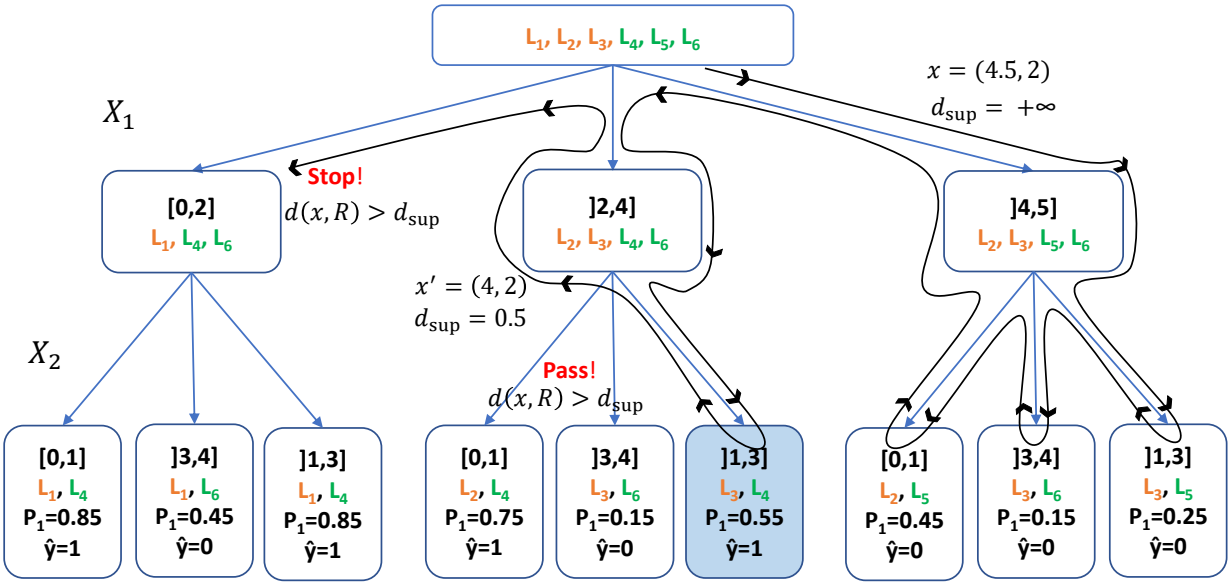

FIGURE 3 – L'arbre de recherche basé sur la Figure 2 et le processus de génération d'exemple contrefactuel pour une requête $x = (4.5, 2)$. L'exemple contrefactuel optimal se trouve dans la région intersectée par L_3 et L_4.

l'exemple contrefactuel d'apprentissage le plus proche peut être grande. Pire encore, lorsque plusieurs caractéristiques protégées (PF) sont considérées, cette approche peut échouer à trouver un exemple contrefactuel initial qui satisfasse les contraintes. Par conséquent, nous proposons une nouvelle stratégie pour trouver un exemple contrefactuel virtuel initial, que nous appelons One-dimensional Change CounterFactual (OCCF). Pour un x donné, OCCF vise à résoudre l'équation (4) avec la contrainte supplémentaire que x et z ne diffèrent que par une seule caractéristique. Ce problème, qui peut être très facilement résolu, a des liens avec la notion d'espérances conditionnelles individuelles [3] qui estiment comment la probabilité (ou la décision) $f(x)$ d'un classifieur varie en fonction d'une modification de x lorsque toutes les autres valeurs sont fixes.

Notons que, dans une forêt aléatoire, pour une requête x, il suffit de considérer un nombre fini de valeurs possibles pour une caractéristique modifiable $\mathscr{X}_d$, définies par les valeurs de coupure obtenues sur tous les arbres pour cette caractéristique. Cependant, il est possible qu'un OCCF n'existe toujours pas lorsque plusieurs caractéristiques sont protégées, bien que cette situation ait été expérimentalement beaucoup moins rencontrée avec OCCF qu'avec l'approche MO. Dans ce cas, certaines contraintes doivent être relaxées en "déprotégeant" certaines caractéristiques non modifiables. Soulignons que ici cette relaxation n'est utilisée que pour l'initialisation de d_{sup} et que notre méthode ne détermine donc pas d'exemple contrefactuel comportant des caractéristiques protégées modifiées.

4 Résultats expérimentaux

4.1 Efficacité de la procédure d'extraction

Dans cette expérience, nous évaluons l'efficacité de la procédure d'extraction d'explications contrefactuelles proposée sur quatre jeux de données. Le nombre d'arbres dans l'ensemble est de 50 pour tous les jeux de données, et la profondeur maximale des arbres est respectivement de 10, 8, 7, et 14. L'efficacité est évaluée selon trois critères : le nombre de régions à explorer après filtrage par différentes approches

TABLEAU 1 – Nombre moyen de feuilles à explorer.

Données	Original	PF	MO	PF+MO	OCCF	PF+OCCF
Compas	7236	2226.86	849.87	732.54	418.36	**305.36**
Heloc	8784	——	5268.12	——	**106.52**	——
Pima	2522	1081.27	1007.43	719.53	133.90	**128.77**
Wine	8949	——	3277.05	——	**761.38**	——

TABLEAU 2 – Distance moyenne entre la requête et l'exemple contrefactuel initial (à gauche), et temps moyen consommé pour la recherche du contrefactuel final (à droite).

	Distance initiale du CF				Temps de recherche du CF (s)			
Données	MO	PF+MO	OCCF	PF+OCCF	MO	PF+MO	OCCF	PF+OCCF
Compas	0.078	0.134	**0.040**	0.058	1.091	0.421	0.580	**0.284**
Heloc	0.273	——	**0.011**	——	4.570	——	**1.274**	——
Pima	0.215	0.273	**0.034**	0.041	5.600	4.991	3.589	**3.277**
Wine	0.192	——	**0.060**	——	5.745	——	**4.667**	——

d'initialisation, la distance entre la requête et l'exemple contrefactuel initial, et le temps nécessaire pour extraire tous les contrefactuels. Notons que Compas et Pima ont respectivement une et quatre caractéristiques protégées, alors qu'aucune caractéristique protégée n'a été considérée pour Heloc et Wine.

Les tableaux 1 et 2 indiquent que l'exploitation des caractéristiques protégées peut contribuer de manière notable à réduire le nombre de régions à explorer, puisqu'elle limite la recherche des contrefactuels à un sous-espace des caractéristiques. L'initialisation de l'OCCF que nous proposons peut générer des contrefactuels initiaux qui sont beaucoup plus proches de la requête x par rapport à la MO : en conséquence, nous pouvons filtrer beaucoup plus de régions, et donc réduire de manière très significative le temps nécessaire pour atteindre une solution.

4.2 Études de cas

Cas 1 : données Pima

L'ensemble de données Pima peut être utilisé pour prédire si un patient est diabétique ou non, sur la base de diverses mesures : le nombre de grossesses (PG) ; le taux de Glucose ; la Pression artérielle (BP) ; l'épaisseur de la peau (ST) ; le taux d'Insuline sérique sur 2 heures (mu U/ml) ; l'indice de masse corporelle (BMI) ; la fonction pedigree du diabète (DPF) ; et l'Age. La classe est $y = 0$ pour un non-diabétique, $y = 1$ pour un diabétique. Ici, l'âge, le nombre de grossesses, les valeurs du DPF et l'épaisseur de la peau sont difficiles à modifier (considérés comme des caractéristiques protégées), tandis que le glucose, l'insuline, BMI et la pression artérielle sont des caractéristiques exploitables (modifiables). Nous avons choisi Pima comme exemple car l'explicabilité peut avoir un intérêt pratique significatif dans le domaine médical.

TABLEAU 3 – Exemples d'explications contrefactuelles tirées de l'ensemble de données Pima.

	PGs	Glucose	BP	ST	Insulin	BMI	DPF	Age
x_1	0	165	90	33	680	52.3	0.427	23
x_1^0	0	**154.5**$\downarrow$	90	33	680	**47.7**$\downarrow$	0.427	23
x_1^1	0	**165.5**$\uparrow$	90	33	680	52.3	0.427	23
x_2	1	122	90	51	220	49.7	0.325	31
x_2^0	1	**121.5**$\downarrow$	90	51	**128**$\downarrow$	**49.05**$\downarrow$	0.325	31
x_2^1	1	**126.5**$\uparrow$	90	51	220	49.7	0.325	31

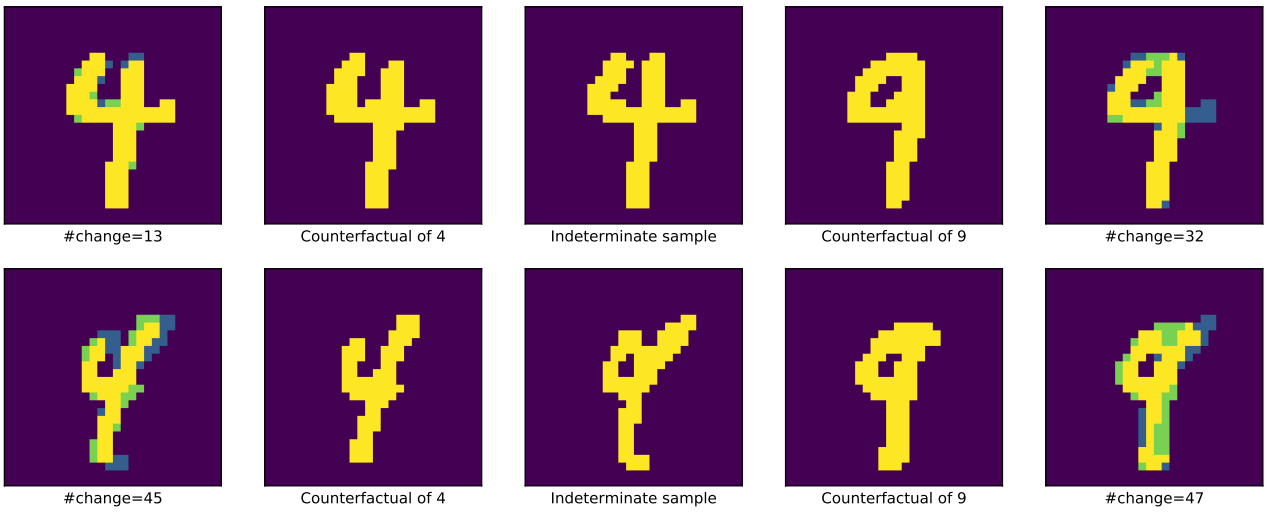

FIGURE 4 – Exemples de chiffres indéterminés (au centre) et de contrefactuels correspondants de la classe 4 (à gauche) et 9 (à droite). Les images les plus à gauche et à droite montrent les pixels à ajouter (vert) et à supprimer (bleu) afin d'obtenir le contrefactuel.

Le Tableau 3, présente deux exemples à titre d'illustration. La requête x_1 correspond à un patient non diabétique. Notons tout d'abord qu'il est proche d'être classé comme diabétique puisque le contrefactuel x_1^1 de cette classe est très proche. Cela montre que la forêt aléatoire prudente peut aider à gérer l'incertitude découlant de la pauvreté des données, en détectant les instances pour lesquelles la décision est incertaine et en fournissant des informations sur leur étiquette réelle. En outre, le contrefactuel non-diabétique x_1^0 suggère une manière possible de maintenir une bonne santé, en réduisant le niveau de BMI et de glucose.

La requête x_2 correspond à un patient diabétique. L'indétermination vient de la caractéristique Glucose, puisqu'on peut obtenir une prédiction correcte (diabétique) en modifiant sa valeur uniquement. En revanche, pour obtenir le contrefactuel non-diabétique x_2^0, une diminution importante du taux d'insuline est nécessaire, ce qui est cohérent avec le fait que les valeurs élevées d'insuline sérique sur 2 heures sont courantes chez les patients diabétiques de type II.

Case 2 : données MNIST

MNIST est une grande base de données de chiffres manuscrits contenant environ 60 000 cas d'entraînement et 10 000 cas de test. Dans notre expérience, les chiffres 4 et 9 ont été sélectionnés et 40 composantes principales ont été extraites des données originales pour entraîner une forêt aléatoire prudente composée de 50 arbres de profondeur 10. Nous avons généré des contrefactuels dont nous avons exigé qu'ils appartiennent à la classe ciblée avec une croyance d'au moins 0,75, afin de garantir que l'instance est crédible après l'application de la transformation PCA inverse. La génération de contrefactuels d'une requête aide à comprendre quelles parties de l'image sont responsables de l'indétermination de la décision. Ce point est illustré à l'aide de deux instances représentées sur la figure 4. Nous pouvons voir comment les deux exemples indéterminés (au centre) devraient être modifiés pour être classés de façon précise, soit comme un "4", soit comme un "9", et que ces modifications ont visuellement du sens.

5 Conclusion

Dans cet article, nous avons proposé une procédure pour extraire des contrefactuels pour des instances indéterminées, c'est-à-dire pour lesquelles aucune décision précise n'a pu être prise, afin d'interpréter et d'expliquer l'indétermination du classifieur. L'algorithme présenté dans cet article est spécifique aux forêts aléatoires prudentes. Il est basé sur un algorithme proposé dans le cas de l'extraction de contrefactuels précis. Nos modifications permettent de filtrer les régions de l'espace des caractéristiques à explorer et de générer le contrefactuel initial plus proche de la requête afin d'accélérer le processus d'extraction.

Cette amélioration de l'efficacité de la procédure, ainsi que l'utilité de notre approche pour expliquer l'indétermination du classifieur, ont été démontrées sur plusieurs expériences. Dans des recherches à venir, nous étudierons comment les explications contrefactuelles peuvent être utilisées pour estimer l'importance des caractéristiques, et pour identifier les régions d'incertitude significative dans l'espace des caractéristiques. Nous prévoyons également d'utiliser les contrefactuels dans un processus d'apprentissage actif pour réduire l'indécision des classifieurs prudents.

Déclaration

Cet article est une traduction française de [11].

Références

[1] P. Blanchart. An exact counterfactual-example-based approach to tree ensemble models interpretability. *arXiv preprint arXiv :210514820*.

[2] L. Breiman. Random forests. *Machine learning*, 45(1) :5–32, 2001.

[3] A. Goldstein, A. Kapelner, J. Bleich and E. Pitkin. Peeking inside the black box : Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, 24(1) :44–65, 2015.

[4] R. Guidotti. Counterfactual explanations and how to find them : literature review and benchmarking. *Data Mining and Knowledge Discovery* pp 1–55, 2022.

[5] C. Molnar. *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. lulu.com, 2020.

[6] G. Shafer. *A mathematical theory of evidence*. Princeton university, 1976.

[7] S. Verma, J. Dickerson, and K. Hines. Counterfactual Explanations for Machine Learning : A Review. *NeurIPS 2020 Workshop : ML Retrospectives, Surveys & Meta-Analyses (ML-RSA)*, 2020.

[8] S. Wachter, B. Mittelstadt B, and C. Russell C. Counterfactual explanations without opening the black box : Automated decisions and the gdpr. *Harvard Journal of Law & Technology* , 31 :841, 2017.

[9] P. Walley. Inferences from Multinomial Data : Learning About a Bag of Marbles. *Journal of the Royal Statistical Society : Series B (Methodological)*, 58 :3–34, 1996.

[10] H. Zhang, B. Quost, and MH. Masson. Cautious Random Forests : a new decision strategy and some experiments. *International Symposium on Imprecise Probability : Theories and Applications*. PMLR, p. 369–372, 2021.

[11] H. Zhang, B. Quost, and MH. Masson. Explaining Cautious Random Forests via Counterfactuals. *To appear in the Proceedings of the 10th International Conference on Soft Methods in Probability and Statistics (SMPS 2022)*.