

Rapport final - Parcours IA

Data Challenge - Biosonar : détection de clics d'odontocètes

Chimbault Thomas, **Gonçalves** Tristan

ENSC - École Nationale Supérieure de Cognitique

E-mails:

- thomas.chimbault@ensc.fr
- tristan.goncalves@ensc.fr

09 Février 2024

I. Introduction

Ce projet s'inscrit dans le cadre du parcours IA de troisième année de l'École Nationale Supérieure de Cognitique. Parmi les propositions de projets disponibles, nous avons décidé de concourir dans un data challenge proposé par l'Université de Toulon sur le site [ChallengeData](#), hébergé par l'ENS.

A. Contexte et objectifs initiaux

L'objectif du data challenge est tout d'abord d'atteindre le benchmark défini pour ce projet, fixé à **0.7509**, afin d'assurer une performance optimale dans la détection de la présence d'un clic d'odontocète dans des enregistrements audios sous-marins. Ces enregistrements proviennent du projet CARI'MAM, qui étudie la biodiversité de l'arc des Antilles et enregistre l'activité sous-marine depuis 2017. Les bouées qui permettent ces enregistrements capturent énormément de données. De ce fait, ces dernières sont difficiles à exploiter, ce qui motive à la création d'un modèle permettant d'aider les chercheurs dans leur démarche. Dans le cadre de ce défi, un deuxième objectif est de concevoir un modèle avec une empreinte environnementale minimale, en réduisant la consommation énergétique lors de son entraînement. Cette considération reste une priorité, en harmonie avec le challenge axé sur la préservation de la faune maritime. Cela se traduit non seulement par la détection des



odontocètes pour mieux comprendre leur présence, mais également par une approche responsable, évitant des répercussions déraisonnables sur l'environnement lors de la création du modèle.

B. Objectifs atteints

À la suite de ce challenge, nous avons réussi à obtenir un modèle léger atteignant des performances de **95.66%** pour la détection de clics d'odontocètes dans des audios bruités. De plus, le contrôle de l'impact environnemental a aussi été maîtrisé avec un entraînement relativement économe de 0.65 kWh. Le modèle est explicité dans les ["Choix techniques - réseaux de convolution 1d"](#). Grâce à cela, nous avons fini quatrième au classement public.

II. Données

A. Description de la base d'entraînement

Le jeu de données comprend 23 168 fichiers audio, catégorisés en enregistrements avec ou sans clic, avec des étiquettes 0 ou 1 respectivement. Près de la moitié des enregistrements (9450) indiquent la présence d'au moins un clic, suggérant une ampleur suffisante pour obtenir des résultats satisfaisants en détection binaire. Les données proviennent de 8 stations des Caraïbes, chacune représentée dans les enregistrements.

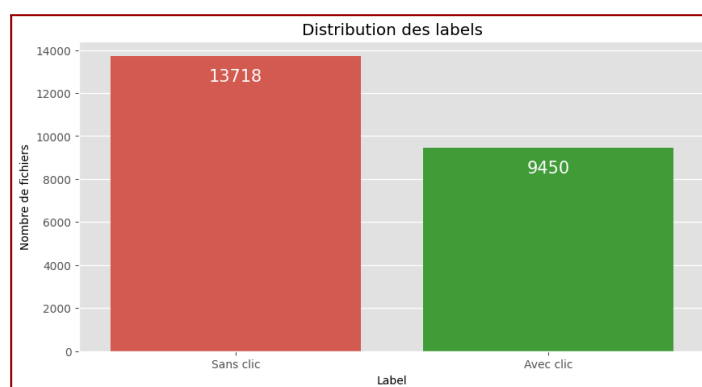


Figure 1 : Distribution et répartition des labels du jeu de données

Les échantillons audio de 200 ms à une fréquence d'échantillonnage de 256 kHz sont illustrés dans la figure 3. Certains enregistrements sans clic sont clairement distinguables, tandis que d'autres peuvent être confondus en raison du bruit ambiant.

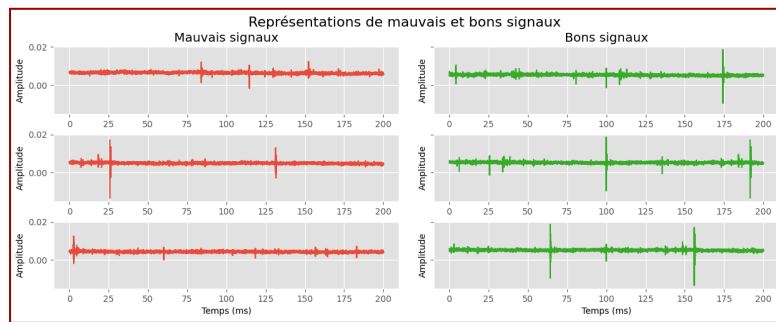


Figure 2 : Représentation de “mauvais” et “bons” enregistrements

B. Description de la base d'évaluation

La base d'évaluation se compose de 950 audios de même durée et fréquence que l'ensemble d'entraînement, mais enregistrés sur 2 stations différentes des Caraïbes. Les résultats sont évalués à l'aide du score ROC AUC, mesurant la capacité du modèle à distinguer les classes positives et négatives, avec un score élevé indiquant de meilleures performances, tandis que 0,5 représente le hasard.

III. Choix techniques

Afin de répondre à la problématique, nous avons eu recours à plusieurs approches, des modèles de scikit-learn aux réseaux de convolution, en passant par les réservoirs. Nous allons voir ces différentes méthodes un peu plus en profondeur.

A. Approches “classiques” de machine learning

Parmi les méthodes classiques de machine learning, nous avons utilisé la Logistic Regression, les Decision Trees, Bagged Trees, Random Forest et XGBClassifier. Ces approches nous ont permis d'obtenir des résultats meilleurs qu'un simple aléatoire, mais ils n'ont pas été convaincants pour autant : le meilleur score obtenu a été de **0.6460**.

B. Reservoir computing

L'approche par reservoir computing, utilisant la librairie ReservoirPy de l'Inria Bordeaux, ne nous a pas permis d'obtenir de bons résultats. En effet, les sorties des réservoirs n'étant pas des probabilités (certains résultats dépassant 1, d'autres étant négatifs), nous avons dû



arrondir ces derniers afin de pouvoir se rapprocher d'un résultat cohérent vis-à-vis de la ROC AUC. Ceci a donc totalement biaisé les soumissions, ce qui nous a conduits au faible score de **0.48**.

L'approche par réservoir a été introduite tardivement dans le projet, lors de la semaine d'AI4Industry, où nous avons appris à utiliser la librairie. Il est possible que notre utilisation dans ce contexte n'ait pas été optimale, entraînant une approche moins efficace de la librairie.

C. Réseaux de convolution

a. Augmentation de données

Avant de nous aventurer vers des modèles plus complexes, nous avons mis en place un processus de Data Augmentation visant à accroître le volume de données pour l'entraînement et ainsi renforcer la robustesse de nos modèles. À cet effet, nous avons introduit un filtre passe-bande d'ordre 6 pour extraire l'ensemble des fréquences audio comprises entre 5 kHz et 100 kHz. Parallèlement, nous avons enrichi la base de données en appliquant des décalages de hauteur et temporels, contribuant à rendre le modèle plus généraliste.

b. Convolution 2D

Les signaux audio étant représentables sous forme de spectrogrammes, nous avons voulu essayer une approche basée sur les réseaux de convolution 2D. Néanmoins, bien que les résultats obtenus avec cette approche aient été satisfaisants, avec un score de **0.86**, nous avons décidé de ne pas poursuivre dans cette direction en raison de la consommation énergétique élevée de l'entraînement du modèle, atteignant 2.01 kWh pour un réseau de petite taille. Nous avons estimé que cette consommation était excessivement élevée et a conduit à reconsidérer notre choix initial.

c. Convolution 1D

Nous avons fait un choix majeur en adoptant un réseau de neurones convolutifs 1D comme architecture principale pour notre projet. Cette décision découle de la capacité du modèle à traiter directement les données audio brutes en entrée. L'architecture du réseau se compose de trois blocs convolutifs, chacun comprenant une couche de convolution avec une fonction d'activation ReLU, suivie d'une couche de MaxPooling, avec des tailles de noyau respectives de 7, 5 et 3. Il est important de souligner que chaque couche convolutive utilise 32 filtres pour extraire des caractéristiques spécifiques des données d'entrée. Par la suite, deux couches Dense ont été ajoutées, comportant respectivement 128 et 64 neurones, chacune appliquant la régularisation L2 avec un coefficient de 0.01 pour atténuer les biais et prévenir tout

sur-apprentissage. Un dropout de 0.5 a été introduit pour renforcer la régularisation et éviter le sur-apprentissage, suivi d'une couche Dense avec un seul neurone et une fonction d'activation sigmoïde pour la sortie binaire. En ce qui concerne l'optimiseur, Adam a été choisi en raison de ses performances éprouvées dans la convergence rapide du modèle. Pour la fonction de perte, nous avons opté pour la binary cross entropy, étant donné que notre tâche relève d'une classification binaire, plus précisément, la détection.

Après un entraînement étendu sur 30 epochs, avec une stratégie d'arrêt prématuré (ou *early stopping*) configurée pour une patience de 3 epochs, le modèle a atteint un score de **0.9566** lors de la soumission. La consommation électrique pendant l'entraînement s'est élevée à 0.65 kWh, équivalent à la consommation moyenne quotidienne d'un congélateur. Les résultats remarquables de ce modèle, conjugués à une dépense énergétique modérée lors de son entraînement, mettent en lumière la pertinence de ce choix technique spécifique pour notre projet.

IV. Gestion de projet

Le projet s'est déroulé de manière itérative, adoptant une approche cyclique basée sur une revue scientifique. Nous avons mis en œuvre différentes architectures et approches de manière régulière. Notre collaboration au sein des mêmes groupes de TD a facilité l'organisation de réunions régulières, formelles et informelles, ainsi que la collaboration synchrone. Nous avons également intégré des idées provenant de nos échanges, permettant d'améliorer le modèle au fil du temps. Cela inclut l'implémentation de papiers scientifiques et la discussion approfondie sur les choix d'hyperparamètres, ainsi que sur les logiques et cheminements de pensée à adopter.

V. Bilan et prise de recul

Ce projet nous a plongés dans le domaine de la Data Science en utilisant des données réelles pour résoudre des problématiques concrètes. Si nous avions disposé de plus de temps, une réflexion approfondie aurait pu être consacrée au choix des fonctions de perte puisque nous sommes restés durant toutes nos expérimentations avec la binary cross entropy...

Suite à ce challenge, Challenge Data nous a contactés pour obtenir nos codes et nous a invités à présenter nos résultats lors de leur université d'été le 18 mai prochain. Cette reconnaissance vient renforcer notre fierté envers notre travail, et nous sommes enchantés d'avoir apporté une contribution significative au projet.
