Contrôle de TP Apprentissage Automatique 1

KASMI Mohamed amine M1 IAFA

Université Toulouse 3 Paul Sabatier

Mon adresse électronique :
mohamed-amine.kasmi@univ-tlse3.fr

Résumé

La classification d'événements sonores joue un rôle crucial dans de nombreuses applications telles que la surveillance environnementale et la reconnaissance automatique de la parole. Dans cette étude, nous comparons les performances de deux architectures de réseaux de neurones, un MLP (Multilayer Perceptron) et un CNN (Convolutional Neural Network). Nous décrivons les données utilisées, les caractéristiques extraites, ainsi que les détails des modèles MLP et CNN, y compris leur architecture et leur processus d'entraînement. En utilisant un protocole expérimental rigoureux, nous évaluons les performances des deux modèles en termes de précision et analysons les résultats obtenus. Nos conclusions mettent en évidence les avantages et les limites de chaque architecture, offrant ainsi des insights précieux pour le choix de modèle dans des applications de classification d'événements sonores.

Mots Clef

Classification, MLP (Multilayer Perceptron), CNN (Convolutional Neural Network) , Performance.

1 Introduction

La classification d'événements sonores est cruciale dans de nombreuses applications telles que la surveillance environnementale, la reconnaissance automatique de la parole, et bien d'autres. La capacité à identifier et à classifier efficacement les sons permet de mieux comprendre notre environnement acoustique et de faciliter la communication entre les systèmes automatisés et les utilisateurs.
Dans ce contexte, les réseaux de neurones, notamment les architectures comme le MLP (Multilayer Perceptron) et le CNN (Convolutional Neural Network), se sont avérés être des outils puissants pour la classification d'événements sonores. Toutefois, il reste à déterminer quelle architecture est la plus efficace pour cette tâche spécifique.
Cette étude vise à comparer les performances d'un MLP et d'un CNN dans la classification d'événements sonores. En évaluant ces deux architectures dans un cadre expérimental rigoureux, nous cherchons à déterminer laquelle offre les meilleures performances en termes de précision et de généralisation à de nouvelles données.

Dans cet article, nous présentons une analyse approfondie des deux architectures de réseaux de neurones, ainsi que des résultats obtenus lors de leur évaluation. Ces résultats seront discutés et interprétés pour fournir des insights précieux sur le choix d'architecture pour la classification d'événements sonores.

2 Description des données

Le corpus est constitué de 10 concepts audio qui sont :
- tronçonneuse (chainsaw)
- tic-tac d'une horloge (clock_tick)
- craquement de feu (crackling_fire)
- pleurs de bébé (crying_baby)
- chien (dog)
- hélicoptère (helicopter)
- pluie (rain)
- coq (rooster)
- bruit des vagues (sea_waves)
- éternuement (sneezing)

Nous voulons classer les différents fichiers audio suivant ces 10 classes.
Les données sont réparties équitablement entre les différentes classes, avec un nombre d'exemples équilibré pour chaque classe. Chaque spectrogramme présente des caractéristiques telles que la résolution temporelle et fréquentielle, qui sont importantes pour l'analyse des événements sonores.

Les caractéristiques extraites des données audio comprennent les spectrogrammes, qui représentent une représentation temps/fréquence du signal sonore. Les spectrogrammes sont obtenus à l'aide de la bibliothèque Librosa, qui permet de convertir les fichiers audio en spectrogrammes à l'aide de la transformation de Mel et de la mise à l'échelle logarithmique des puissances.

Ce prétraitement des données audio permet d'obtenir des représentations exploitables par les modèles de classification, en capturant les informations importantes sur la structure temporelle et fréquentielle des événements sonores.

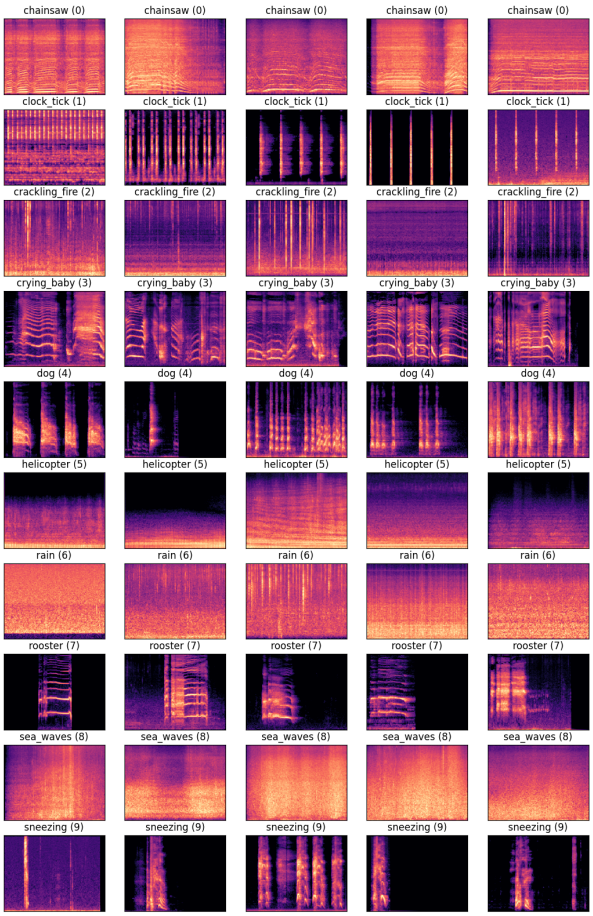

FIGURE 1 – spectrogrammes

3 Méthodes

3.1 Modèle MLP

Pour le modèle MLP (Multilayer Perceptron), nous avons utilisé une architecture simple avec une seule couche cachée. Voici les détails de l'architecture :
- **Nombre de couches :** 2 (1 couche cachée + 1 couche de sortie)
- **Nombre de neurones par couche :**
Couche cachée : 50 neurones
Couche de sortie : 10 neurones (correspondant aux 10 classes d'événements sonores)

- **Fonction d'activation :** ReLU (Rectified Linear Unit) pour la couche cachée et linéaire pour la couche de sortie.

- **Prétraitement des données pour le MLP :**
Les données ont été prétraitées en convertissant les fichiers audio en spectrogrammes à l'aide de la bibliothèque Librosa.
Les spectrogrammes ont été normalisés et redimensionnés pour correspondre aux exigences d'entrée du MLP.
- **Optimisation et entraînement du modèle :**
Nous avons utilisé la fonction de perte de l'entropie croisée (Cross Entropy Loss) comme fonction de perte.
L'optimiseur Adam a été utilisé avec un taux d'apprentissage de 0.0001.
Le modèle a été entraîné sur 10 epochs avec une taille de lot (batch size) de 32 exemples.

3.2 Modèle CNN

Pour le modèle CNN (Convolutional Neural Network), nous avons utilisé une architecture typique pour la classification d'images. Voici les détails de l'architecture :
- **Architecture du CNN :**
3 couches de convolution avec des noyaux de taille 3x3 et 8, 16 et 32 canaux de sortie respectivement.
3 couches de pooling (MaxPooling) avec une taille de fenêtre de 2x2 et un stride de 2.
2 couches fully-connected (entièrement connectées) avec 50 neurones pour la première couche et 10 neurones pour la couche de sortie.
- **Prétraitement des données pour le CNN :**
Les mêmes données de spectrogramme que pour le MLP ont été utilisées en entrée du CNN.
- **Optimisation et entraînement du modèle :**
La même fonction de perte (entropie croisée) et le même optimiseur (Adam) que pour le MLP ont été utilisés.
Le CNN a également été entraîné sur 10 epochs avec une taille de lot de 32 exemples.

Ces méthodes ont été mises en œuvre et évaluées pour comparer les performances des deux architectures (MLP et CNN) dans la classification d'événements sonores.

4 Protocole expérimental

Dans cette section, nous décrirons en détail le protocole expérimental que nous avons suivi pour comparer les performances des modèles MLP et CNN dans la classification d'événements sonores.

4.1 Division des données en ensembles d'entraînement et de test

Nous avons divisé les données en deux ensembles distincts : un ensemble d'entraînement et un ensemble de test. L'ensemble d'entraînement a été utilisé pour entraîner les modèles MLP et CNN, tandis que l'ensemble de test a été réservé pour évaluer les performances des modèles sur des données non vues.

4.2 Paramètres d'apprentissage utilisés pour chaque modèle

Pour chaque modèle, nous avons défini les paramètres d'apprentissage suivants :
Modèle MLP :
- Taux d'apprentissage : 0.0001
- Nombre d'epochs : 10
- Taille de lot (batch size) : 32

Modèle CNN :
- Taux d'apprentissage : 0.0001
- Nombre d'epochs : 10
- Taille de lot (batch size) : 32

4.3 Méthodes d'évaluation des performances

Pour évaluer les performances des modèles, nous avons utilisé les mesures suivantes :
Précision (Accuracy) : Il s'agit du pourcentage de prédictions correctes faites par le modèle sur l'ensemble de test.
Courbes d'apprentissage : Nous avons tracé les courbes d'apprentissage pour visualiser l'évolution de la perte et de la précision au cours de l'entraînement.

4.4 Améliorations potentielles apportées aux modèles

Pour améliorer les performances des modèles, plusieurs techniques peuvent être explorées, notamment :
Augmentation des données : En appliquant des transformations aux données d'entraînement pour augmenter la diversité des exemples.
Réglage des hyperparamètres : En ajustant les hyperparamètres tels que le taux d'apprentissage, le nombre d'epochs et la taille de lot pour optimiser les performances du modèle.
Utilisation de techniques de régularisation : Comme la régularisation L1/L2 pour réduire le surapprentissage.
Exploration d'architectures plus complexes : En testant des architectures de réseaux neuronaux plus avancées pour capturer des caractéristiques plus fines dans les données.

5 Résultats

Dans cette section, nous présenterons les résultats obtenus pour chaque modèle (MLP et CNN) dans la classification d'événements sonores.

5.1 Précision des modèles

Les résultats de précision obtenus pour chaque modèle sont les suivants :
MLP :
- Précision finale Train : 0.19375
- Précision finale Test : 0.1625
CNN :
- Précision finale Train : 0.78125
- Précision finale Test : 0.6875

5.2 Courbes d'apprentissage

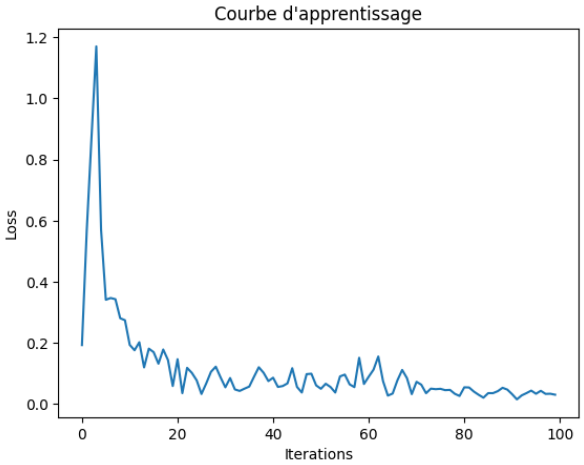

FIGURE 2 – Courbe MLP

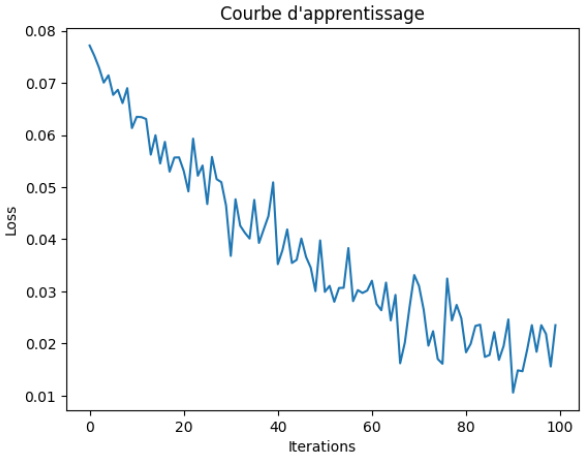

FIGURE 3 – Courbe CNN

5.3 Interpretation

Le modèle CNN a démontré une meilleure précision que le modèle MLP. À partir de la courbe et les résultats de precision, on peut voir que ces résultats suggèrent que le CNN a mieux appris les caractéristiques des données .

6 Interprétation des résultats des matrices de confusion

Pour interpréter les résultats des matrices de confusion pour les modèles MLP et CNN, nous devons comprendre comment elles sont structurées.
Chaque matrice de confusion est une grille où les lignes représentent les vraies classes et les colonnes représentent les classes prédites par le modèle. Chaque cellule de la matrice contient le nombre d'échantillons qui appartiennent à une certaine vraie classe et qui ont été prédits comme appartenant à une certaine classe prédite.
Examinons la première matrice de confusion pour le modèle MLP :

Interpretation :
- La première ligne montre que le modèle a correctement prédit 6 échantillons appartenant à la classe 0 et a prédit à

```
tensor([[3., 0., 0., 0., 4., 0., 0., 0., 1., 0.],
        [0., 4., 0., 0., 0., 4., 0., 0., 0., 0.],
        [0., 6., 2., 0., 0., 0., 0., 0., 0., 0.],
        [0., 2., 0., 2., 4., 0., 0., 0., 0., 0.],
        [0., 2., 0., 1., 4., 1., 0., 0., 0., 0.],
        [2., 2., 0., 0., 2., 0., 2., 0., 0., 0.],
        [0., 1., 0., 2., 0., 1., 0., 0., 4., 0.],
        [0., 0., 0., 1., 4., 0., 2., 1., 0., 0.],
        [0., 0., 0., 1., 6., 0., 0., 0., 1., 0.],
        [0., 3., 0., 0., 1., 0., 0., 0., 0., 4.]])
```

FIGURE 4 – Matrice de confusion MLP

tort 2 échantillons de la classe 8.
- La deuxième ligne montre que le modèle a correctement prédit 6 échantillons appartenant à la classe 1 et 2 échantillons appartenant à la classe 2, mais n'a rien prédit pour les autres classes. Et ainsi de suite pour les autres lignes.

```
tensor([[6., 0., 0., 0., 0., 0., 0., 0., 2., 0.],
        [0., 6., 2., 0., 0., 0., 0., 0., 0., 0.],
        [0., 1., 7., 0., 0., 0., 0., 0., 0., 0.],
        [0., 1., 0., 6., 1., 0., 0., 0., 0., 0.],
        [0., 4., 0., 0., 3., 0., 0., 1., 0., 0.],
        [2., 0., 0., 0., 0., 2., 3., 0., 1., 0.],
        [5., 0., 0., 0., 0., 1., 1., 0., 1., 0.],
        [0., 0., 0., 0., 0., 0., 0., 6., 0., 2.],
        [1., 0., 0., 0., 0., 0., 0., 0., 7., 0.],
        [0., 1., 0., 0., 0., 0., 0., 0., 0., 7.]])
```

FIGURE 5 – Matrice de confusion CNN

Interpretation :
- La première ligne montre que le modèle a correctement prédit 6 échantillons appartenant à la classe 0 et a prédit à tort 2 échantillons de la classe 8.
- La deuxième ligne montre que le modèle a correctement prédit 6 échantillons appartenant à la classe 1 et 2 échantillons appartenant à la classe 2, mais n'a rien prédit pour les autres classes. Et ainsi de suite pour les autres lignes.

En examinant les matrices de confusion pour les modèles MLP et CNN, voici ce qu'on peut observer :
Pour le modèle MLP : Il a des difficultés à distinguer certaines classes, par exemple, il confond souvent la classe 0 avec la classe 4.
Certains échantillons sont correctement classés, mais il y a aussi beaucoup de prédictions incorrectes, comme en témoignent les nombres significatifs hors de la diagonale principale de la matrice.
Pour le modèle CNN : Globalement, il semble mieux performer que le modèle MLP car il a moins de prédictions erronées.
Il réussit souvent à classifier correctement les échantillons, comme en témoigne le nombre élevé d'échantillons correctement classés sur la diagonale principale de la matrice.
En résumé, le modèle CNN semble mieux performer que le modèle MLP en termes de classification précise des différentes classes.

7 Conclusion

Les résultats de cette étude comparative démontrent que le modèle CNN surpasse le modèle MLP dans la classification des événements sonores, grâce à sa capacité à capturer les caractéristiques spatiales et temporelles des spectrogrammes. Cela souligne l'importance de choisir des architectures de réseau adaptées à la nature des données. Pour des applications de classification d'événements sonores exigeantes en précision, l'utilisation de CNN est fortement recommandée. Cependant, des efforts de recherche supplémentaires pourraient se concentrer sur l'optimisation des hyperparamètres et l'exploration de nouvelles architectures CNN pour améliorer encore les performances.

Références

[1] https ://datascientest.com/convolutional-neural-network
[2] https ://fr.wikipedia.org/wiki/Perceptron_multicouche
[3] https ://medium.com/data-science-bootcamp/multilayer-perceptron-mlp-vs-convolutional-neural-network-in-deep-learning-c890f487a8f1
[4] https ://neuroconnection.eu/quand-utiliser-les-reseaux-de-neurones-mlp-cnn-et-rnn/