

Résumé de la thèse : « Optimisation des réseaux de neurones événementiels pour la fusion de capteurs neuromorphiques »

Alors que les applications de l'IA continuent de se développer dans divers domaines, la demande de méthodes informatiques plus efficaces et plus durables s'est intensifiée.

En réponse aux préoccupations concernant la consommation d'énergie des modèles d'IA traditionnels, les réseaux de neurones à impulsions (SNN) sont particulièrement attrayants en raison de leur efficacité énergétique, car ils imitent le fonctionnement du cerveau avec des impulsions discrètes.

Les capteurs neuromorphiques sont bio-inspirés et traitent les informations de la même manière que le cerveau humain, ce qui leur permet de traiter efficacement des données discrètes. Ces capteurs excellent dans la capture et le traitement des données uniquement lorsque des changements dans l'environnement ou des événements importants se produisent, ce qui permet d'économiser de l'énergie en évitant le traitement continu de gros volumes de données.

Ma thèse vise à optimiser la consommation d'énergie en minimisant le nombre d'opérations et de calculs requis dans les réseaux de neurones pour traiter les données fournies par la fusion de capteurs. Pour cela, nous utilisons des réseaux de neurones à impulsions qui ont une bonne performance pour un faible coût énergétique.

En outre, le réseau simulé en software devra être reproductible en hardware pour des systèmes embarqués.

Nous travaillons actuellement sur la TIMIT, une base de données couramment utilisée pour la classification audio. Il s'agit d'enregistrements audio transcrits phonétiquement et annotés, qui fournissent des données précieuses pour l'étude de la reconnaissance vocale.

La première approche pour traiter les données issues de la TIMIT est une approche classique et consiste à pré-traiter l'audio en cepstrogrammes de coefficients cepstraux de fréquence Mel (MFCC). Cela consiste à découper l'audio en sous-échantillons, à les convertir en une représentation fréquentielle, à appliquer un filtrage à l'échelle de Mel pour imiter la perception auditive humaine et, enfin, à en extraire les MFCC.

Les réseaux de neurones apprendront à associer les motifs des MFCC à des phonèmes spécifiques. Cela permet au réseau de neurones de reconnaître et de classer les phonèmes sur la base des motifs et des caractéristiques extraits des "images" MFCC du signal audio.

La seconde approche est plus proche du fonctionnement d'un capteur neuromorphique : l'audio est traité à travers différents filtres passe-bande, dont les paramètres dépendent du spectre de fréquence du signal. La conversion en impulsions est ensuite effectuée en envoyant les sorties rectifiées des filtres aux neurones à impulsions.

Ensuite, pour chaque approche, un réseau de neurones traditionnel et un réseau de neurones à impulsions avec les mêmes architectures sont appliqués. Nous comparons la performance et le coût énergétique des quatre résultats.

Ce qui est fait pour la base de données TIMIT sera fait pour d'autres bases de données, puis la fusion de différentes bases de données sera explorée.

Par exemple, une base de données avec l'image du trafic et une base de données avec le son du trafic.

Valentin MEUNIER

Doctorant - Université de Bordeaux

Laboratoire IMS