

Titre : Réseau neuronal pour l'acheminement global centré sur la capacité d'acheminement de l'automatisation de conception de circuits électroniques (EDA)

Conférenciers : André Ivanov, PhD et Sebastian Zhou

Résumé

L'acheminement des fils de connexion pour les systèmes sur puce complexes comporte des difficultés computationnelles redoutables et représente l'un des problèmes les plus épineux de la soi-disant conception physique en amont des systèmes sur puce (circuits intégrés). La congestion engendrée, un facteur clé de la capacité d'acheminement, augmente habituellement le nombre de voies et de détours dans l'agencement, ce qui affecte généralement de manière négative la performance globale du circuit. De plus, les zones congestionnées dans l'agencement peuvent entraîner la création de courts-circuits et de ruptures dans l'interconnexion, réduisant ainsi le rendement de fabrication ou la fiabilité des systèmes sur puce.

Les approches traditionnelles utilisent des algorithmes pour produire des modèles fixes d'estimation de la congestion et s'adaptent mal aux nœuds technologiques évoluant vers des dimensions plus réduites. Nos travaux portent sur l'exploration et le développement d'un réseau neuronal de régression pour la reconstruction d'images qui peut prédire une image de la congestion en fonction des propriétés de l'agencement du circuit. De plus, pour résoudre le problème de compatibilité de format engendré par les différents outils d'automatisation de conception de circuits électroniques au cours de la génération des caractéristiques d'entrée, nous avons mis au point un algorithme de « pont » pour construire des images de caractéristiques universelles, sans égard aux outils choisis. Nous obtenons une qualité d'acheminement global comparable, qui résulte en moyenne en 20 % moins de débordements initiaux. Cette amélioration initiale de la qualité est reflétée dans la solution d'acheminement final, alors que d'autres techniques d'estimation demandent jusqu'à 5 % plus d'itérations d'acheminement. En outre, l'acheminement de toutes les références d'étalonnage, au total, est au moins trois fois plus rapide.

Biographie (André)

M. André Ivanov est professeur de génie électrique et informatique à l'Université de la Colombie-Britannique. Il a publié de nombreux articles sur plusieurs sujets de recherche associés à la conception et à la mise à l'essai de systèmes sur puce. Il détient également de nombreux brevets. Il est un membre Golden Core de la Computer Society, un Fellow de l'IEEE, un Fellow de l'Académie canadienne du génie, un Fellow de l'Institut canadien des ingénieurs et un ingénieur professionnel en Colombie-Britannique. Les champs d'intérêt de recherche actuels de M. Ivanov se concentrent sur de nouvelles solutions et méthodologies afin d'alléger les problèmes de fiabilité associés aux systèmes sur puce dans les applications de l'Internet des objets (IdO). Ces méthodologies comprennent les simulations dynamiques moléculaires combinées à des approches d'apprentissage machine. M^r Ivanov mène également des recherches sur les applications de l'apprentissage machine à l'automatisation de conception de circuits électroniques (EDA) des systèmes sur puce.

Biographie (Sebastian)

Zhonghua (Sebastian) Zhou est doctorant au département de génie électrique et informatique à l'Université de la Colombie-Britannique. Il compte une année et demie d'expérience de travail en automatisation de conception de circuits électroniques (EDA) dans une entreprise en démarrage de ce domaine industriel de même que chez Synopsis. Ses travaux de recherche au doctorat portent sur l'optimisation des techniques d'acheminement global de systèmes sur puce. Le projet actuel se concentre sur l'élaboration de stratégies d'acheminement centrées sur la capacité d'acheminement. Ces stratégies comprennent des estimations de la congestion basées sur l'apprentissage profond et un algorithme d'acheminement qui tient compte de la congestion.

Titre : Optimisation séquentielle probabiliste à plusieurs objectifs de réseaux de neurones à convolution

Présentateurs : Brett Meyer, PhD

Résumé

Avec l'arrivée de réseaux de neurones à convolution (RNC) plus profonds, plus gros et plus complexes, la conception manuelle est devenue une tâche pénible, surtout lorsque le logiciel d'apprentissage machine doit être optimisé pour la performance matérielle. L'optimisation par modèles séquentiels est une méthode efficace pour l'optimisation des hyperparamètres des algorithmes d'apprentissage machine comptant de nombreux paramètres. Elle est en mesure de trouver de bonnes configurations avec un nombre limité d'évaluations en prédisant la performance des candidats avant leur évaluation. L'erreur de prédiction, par contre, mène à des calculs superflus pour entraîner des RNC peu performants. Afin de régler ce problème, nous proposons l'optimisation par modèles séquentiels, qui sélectionne des candidats en fonction d'une estimation probabiliste de leur efficacité de Pareto. Nous utilisons un réseau neuronal pour prédire l'exactitude du modèle étant donné les hyperparamètres du modèle. De plus, nous utilisons l'erreur de prédiction, accompagnée de l'incertitude sur la mesure de la latence d'inférence, afin de quantifier la qualité d'un modèle candidat de deux manières : la probabilité qu'il soit un optimum de Pareto et le nombre attendu de solutions optimales de Pareto actuelles qu'il dominera. Nous évaluons notre méthode proposée sur quatre problèmes de classification d'images. Comparée à une approche déterministe, l'optimisation probabiliste par modèles séquentiels génère constamment des solutions optimales de Pareto qui offrent un meilleur rendement et qui sont concurrentielles par rapport aux modèles de RNC efficaces de pointe. Cela permet une accélération incroyable de la latence d'inférence tout en conservant une exactitude comparable.

Biographie

M. Brett H. Meyer est professeur agrégé au département de génie électrique et informatique à l'Université McGill. Il s'est joint à cette institution en 2011 après plus d'un an comme chercheur postdoctoral à la University of Virginia. Il a reçu son PhD de l'Université Carnegie Mellon en 2009. M. Meyer a publié plus de 50 articles en lien avec ses différents champs

d'intérêt de recherche. Ses travaux portent en général sur les architectures et les algorithmes de conception pour les systèmes embarqués. Toutefois, ses recherches récentes se concentrent sur la sécurité des systèmes embarqués et l'apprentissage machine.