جامعة الموصل
كلية الهندسة

تسريع تدريب الشبكات العصبية الاصطناعية باستخدام تقنيات التوازي

نور موفق جبر الليلة

رسالة ماجستير

علوم في هندسة الحاسوب

بإشراف

الأستاذ المساعد

د. شفاء عبد الرحمن داؤود

1436هـ 2015

Abstract

Training the artificial Neural Network with large number of pattern takes a long time, in this thesis is try to speed up the training of networks using parallelism techniques, including Multicore Central Processing Unit (Multi-Core CPU) and using General Purpose Graphics Processing Unit (GPGPU).

The neural network training in this thesis has been implemented on the six sets of different numbers of data pattern on different sizes neural networks of Multilayer Perceptron neural network (MLP), the speed up was obtained in training artificial neural networks and reduce networks training time in medium and large numbers of data pattern and training speed linearly increased with increasing the number of cores in CPUs, also high training speed was obtained when using Graphical Processing Unit(GPU), while the observed increase in time training with small numbers of data pattern because the cost of communication between the cores which overwhelm the network increased training speed when performing training using the multi-core CPU.

The implementation of neural network training was done using several software environments: MATLAB, CUDA and Open MP were obtained good results in reduced training networking time, the high speed up is obtained when using multi-core CPU and GPU compared with the implementation of the training using a single core CPU,

Where a good speed obtained when performing training parallelism using MATLAB software environment up to × 4 when using 8- cores CPU, Open MP dealing with multi-core CPU where amounted to speed up × 8 compared with the implementation of the training using a single-core CPU using the same software environment, also getting high train speed when performing training proposed networks using GPU reaching × 139 when using GPU with 48- cores and amounted × 304 when used GPU with 384 - core, the highest speed obtained when

implemented networks training by using CUDA software environment reaching to × 743 when using GPU with 384 - cores.

Then, implemented roof line model to calculate performance in processing unit and memory used in the implementation of neural networks training to evaluate the bottlenecks placements in systems.

University of Mosul
College of Engineering

Training Acceleration of Artificial Neural Network using Parallel Techniques

By

Noor Mowafq Gabir Al Layla

Master Thesis In
Computer Engineering

Supervised By
Assistant Professor

Dr. Shefa Abdul Rahman Dawwd

2015 A.D. **1436 A.H.**

المستخلص بغلة الرسالة

تستغرق عملية تدريب الشبكات العصبية الإصطناعية ذات الأعداد الكبيرة من نماذج التدريب وقتا طويلا, في هذه الرسالة محاولة لتسريع تدريب الشبكات العصبية الإصطناعية باستخدام تقنيات التوازي ومنها وحدة المعالجة المركزية متعددة النوى Multicore Central Processing Unit (CPU) و باستخدام وحدة المعالجة الرسومية للأغراض العامة General Purpose Graphics Processing Unit (GPGPU) . إن تنفيذ التدريب للشبكات العصبية الإصطناعية في هذه الرسالة تم على ستة مجموعات من البيانات ذات أعداد مختلفة من النماذج وعلى شبكات عصبية مختلفة الأحجام من نوع شبكة عصبية متعددة الطبقات Multilayer Perceptron (MLP) , تم الحصول على تسريع في تدريب الشبكات العصبية ذات الأعداد المتوسطة والكبيرة لعدد النماذج وكانت سرعة التدريب تزداد مع زيادة عدد النوى في وحدات المعالجة المركزية , وكذلك الحصول على سرعة تدريب عالية عند استخدام وحدة المعالجة الرسومية , بينما لوحظ زيادة زمن تدريب الشبكات ذات الأعداد الصغيرة للنماذج بسبب كلفة الاتصال بين النوى التي تطغى على زيادة سرعة تدريب الشبكة عند تنفيذ التدريب باستخدام وحدة معالجة متعددة النوى بدل وحدة معالجة وحيدة النواة. نفذ تدريب الشبكات المقترحة باستخدام عدة بيئات برمجية : MATLAB , CUDA و Open MP وتم الحصول على نتائج جيدة في تقليل زمن تدريب الشبكات , كانت أعلى سرعة تم الحصول عليها عند استخدام وحدة معالجة الرسوم مقارنة مع تنفيذ التدريب باستخدام وحدة معالجة وحيدة النواة أو وحدة معالجة متعددة النوى إذ تم الحصول على تسريع جيد عند تنفيذ تدريب التوازي باستخدام بيئة MATLAB البرمجية وصلت لغاية ×7 عند استخدام وحدة معالجة ذات ثمان نوى , تم الحصول على تسريع لتدريب الشبكات المقترحة عند تنفيذ الشبكة باستخدام البيئة البرمجية Open MP التي تتعامل مع وحدات المعالجة متعددة النوى حيث بلغ التسارع ×4 مقارنة مع تنفيذ التدريب باستخدام وحدة معالجة وحيدة النواة باستخدام نفس البيئة البرمجية, ايضا تم الحصول على تسريع عالي عند تنفيذ تدريب الشبكات المقترحة باستخدام وحدة معالجة الرسوم التي بلغت ×139 عند استخدام وحدة معالجة الرسوم ذات 48 نواة وبلغت ×304 عند استخدام وحدة معالجة الرسوم ذات 384 نواة, كان أعلى تسارع حصل عليه عند تنفيذ تدريب الشبكات على وحدة معالجة الرسوم باستخدام البيئة البرمجية CUDA إذ بلغت ×743 عند استخدام وحدة معالجة الرسوم ذات 384 نواة. تم تنفيذ نموذج خط السقف (Roofline Model) لبيان أداء وحدات المعالجة والذاكرة المستخدمة في تنفيذ الشبكات العصبية ولبيان مواضع الاختناقات (Bottlenecks) وتقييم الأداء في الأنظمة.

مسؤول الدراسات العليا

د.عمر موفق اليوسف