ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Τσίρμπας Ραφαήλ
Μεταπτυχιακός Φοιτητής

Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

Επόπτης Μεταπτυχιακής Εργασίας: Καθηγητής, Ε. Μαρκάτος

Καθηγητής, Σ. Ιωαννίδης (Επιβλέπων)

Παρασκευή , 6 Νοεμβρίου 2020 , ώρα 14:00 μ.μ.

Διεύθυνση μετάδοσης (url): http://video.ucnet.uoc.gr/live/show/332

Κανάλι YouTube του Τμήματος
https://www.youtube.com/channel/UC7uE3QiMTQjkrpByB_Gnt6Q/live

“ Αποδοτική πρόβλεψη μοντέλων Μηχανικής Μάθησης σε ετερογενή συστήματα ”

Περίληψη

Τα ετερογενή υπολογιστικά συστήματα απαρτίζονται από ένα σύνολο υπολογιστικών συσκευών, κάθε μια από τις οποίες έχει τα δικά της χαρακτηριστικά κατανάλωσης ενέργειας καθώς και την απόδοση της. Ακόμα και σήμερα όμως η πλειοψηφία των εφαρμογών μηχανικής μάθησης χρησιμοποιεί μία μόνο υπολογιστική συσκευή (όπως τον επεξεργαστή ή κάποιον επιταχυντή), για να κάνει προβλέψεις, αφήνοντας τις υπόλοιπες υπολογιστικές συσκευές αδρανείς και ανεκμετάλλευτες. Σε αυτή τη δουλειά,

προτείνουμε μια διαφορετική προσέγγιση στην οργάνωση και στην ανάθεση των προβλέψεων μοντέλων μηχανικής μάθησης σε ετερογενείς συσκευές. Ο αλγόριθμος που υλοποιεί την ανάθεση των εργασιών στις κατάλληλες συσκευές είναι ικανός να ανταποκριθεί γρήγορα στις δυναμικές διακυμάνσεις πραγματικού χρόνου όπως για παράδειγμα, αυξομειώσεις στην είσοδο του συστήματος, υπερφόρτωση εφαρμογών, και αλλαγές στο υπολογιστικό σύστημα. Τα αποτελέσματα της έρευνας μας δείχνουν ότι ο αλγόριθμος μας είναι ικανός να φτάσει τα μέγιστα ποσοστά απόδοσης ανάμεσα σε διαφορετικά μοντέλα μηχανικής μάθησης, προβλέποντας σωστά την κατάλληλη συσκευή με ποσοστό 92.5%, καταναλώνοντας έως και 10% λιγότερη ενέργεια.

University of Crete

Computer Science Department

M.Sc. Thesis presentation / examination

Tsirmpas Rafail

Master's Thesis Supervisor: Professor, E. Markatos

Prof., S. Ioannidis (Thesis CO-Advisor)

Friday, 6 November 2020, 14:00 p.m

Teleconference (will use the e: Presence system), Computer Science Department, University of Crete

(url) : http://video.ucnet.uoc.gr/live/show/332

YouTube channel :

https://www.youtube.com/channel/UC7uE3QiMTQjkrpByB_Gnt6Q/live

“The Best of Many Worlds: Efficient Machine Learning Inference on Heterogeneous Hardware Architectures”

Abstract

Heterogeneous and asymmetric computing systems are composed by a set of different processing units, each with its own unique performance and energy characteristics. Still, the majority of current machine learning applications targets only a single device (the CPU or some accelerator), leaving the rest processing resources unused and idle. In this work, we propose an adaptive scheduling approach that supports heterogeneous and asymmetric hardware, tailored for a diversified set of machine learning models. Our scheduler can respond quickly to dynamic performance fluctuations that occur at real-time, such as data bursts, application overloads and system changes. The experimental results show that it is able to match the peak throughput of a diverse set of machine learning models, by predicting correctly the appropriate device with an accuracy of 92.5%, while consuming up to 10% less energy.