

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ**

**ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ**

**Άσπρη Μαρία  
Μεταπτυχιακή Φοιτήτρια**

**Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης  
Επόπτης Μεταπτ. Εργασίας: Καθηγητής, Π. Τσακαλίδης**

**Πέμπτη, 30/05/2019, 12:00**

**Αίθουσα K206, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης**

**“ Αρχιτεκτονικές Κατανεμημένης Εμβριθούς Μάθησης για Συστάδες Μηχανημάτων  
Περιορισμένων Πόρων ”**

#### **ΠΕΡΙΛΗΨΗ**

Τα τελευταία χρόνια, η Εμβριθής Μάθηση έχει αναχθεί σε δομικό συστατικό για πάρα πολλές υπολογιστικές εφαρμογές, έχοντας καταφέρει να αυτοματοποιήσει την διαδικασία παραγωγής γνωρισμάτων και συνεπώς να εκτοπίσει άλλες, πιο συμβατικές τεχνικές Μηχανικής Μάθησης. Για να είναι αποδοτική, η Εμβριθής Μάθηση πρέπει να έχει πρόσβαση τόσο σε δεδομένα μεγάλου όγκου, όσο και σε μηχανήματα με μεγάλα αποθέματα υπολογιστικής ισχύς και μνήμης. Παράλληλα, τα μηχανήματα χαμηλών υπολογιστικών πόρων είναι ευρέως διαδεδομένα τόσο στην έρευνα όσο και στη βιομηχανία, κυρίως λόγω της άμεσης διαθεσιμότητας τους αλλά και του χαμηλού κόστους τους. Όμως τέτοια υπολογιστικά συστήματα δεν έχουν ούτε τον χώρο που απαιτείται για την αποθήκευση δεδομένων μεγάλου όγκου, ούτε τα αποθέματα μνήμης για να εκτελέσουν αποτελεσματικά εμβριθή μοντέλα. Μια πρόταση για την αντιμετώπιση αυτού του προβλήματος, είναι η μεταφορά αυτών των μοντέλων σε κατανεμημένα περιβάλλοντα, με την ομαδοποίηση μηχανημάτων περιορισμένων πόρων σε Συστάδες.

Ο βασικός στόχος αυτής της μεταπτυχιακής εργασίας είναι η παρουσίαση και η εφαρμογή τεχνικών Κατανεμημένης Εμβριθούς Μάθησης, σε επίπεδο τόσο δεδομένων όσο και αρχιτεκτονικών, με σκοπό να εκμεταλλευτούμε στο μέγιστο βαθμό τους διαθέσιμους πόρους μιας Συστάδας, αλλά και να αξιοποιήσουμε την επίδραση που μπορεί να έχει ένα τέτοιο κατανεμημένο περιβάλλον στην συμπεριφορά συνελικτικών μοντέλων.

Παρουσιάζουμε τα αποτελέσματά της μελέτης μας με την μορφή περιπτωσιολογικών μελετών σε δύο σημαντικά ερευνητικά πεδία, εκείνο της κοσμολογίας και αυτό της τηλεπισκόπησης, χρησιμοποιώντας μια Συστάδα πέντε μηχανημάτων με περιορισμένους πόρους. Στην πρώτη μελέτη, χρησιμοποιούμε τεχνικές κατανομής δεδομένων ώστε να εξετάσουμε την απόδοση ενός συνελικτικού νευρωνικού δικτύου που χρησιμοποιείται για εκτίμηση ερυθρής μετατόπισης, ως προς τον αριθμό των μηχανημάτων μιας Συστάδας αλλά και ως προς τον τρόπο που μοιράζονται σε αυτά τα δεδομένα. Στην δεύτερη μελέτη, προσπαθούμε να σπάσουμε ένα διακλαδωμένο συνελικτικό μοντέλο με πολλές παραμέτρους μεταξύ των μηχανημάτων της Συστάδας, με σκοπό την ταξινόμηση κάλυψης γης. Στόχος μας μέσω της εφαρμογής αυτής είναι να πετύχουμε βέλτιστο διαμοιρασμό φόρτου υπολογιστικής ισχύς, αλλά και να ελαττώσουμε την κίνηση του δικτύου που προκύπτει λόγω της συχνής μεταφοράς δεδομένων μεταξύ των υπολογιστικών συστημάτων.

**Aspri Maria**  
**M.Sc. Thesis**

**Computer Science Department**  
**University of Crete**  
**Master's Thesis Supervisor: Professor, P. Tsakalides**

**Thursday, 30/05/2019, 12:00**  
**Room K206, Computer Science Dept., University of Crete**

**“Architectures of Distributed Deep Learning on Commodity Clusters”**

### **ABSTRACT**

For the last few years, Deep Learning, is becoming an important tool in many computational applications, having trivialized the whole pipeline of feature extraction

and, as a result, replacing other popular Machine Learning algorithms. For Deep Learning to be effective though, it needs not only access to vast amounts of data, but also devices with high computational performance. At the same time, commodity computers, with their high availability and low cost, are a popular choice of hardware and thus widely used, both in industry as well as in academia. However, they lack not only the required space for storing large volume datasets, but also the computational capacity to make Deep Learning a viable choice. In order to address this issue, the solution of commodity clusters was proposed.

The main goal of the present thesis is the study and application of Distributed Deep Learning techniques, through the scope of both data and model parallelization, aiming to effectively migrate Deep Learning on commodity hardware. Our objective is the best possible management of the available Cluster resources, as well as to study and exploit the impact of distributed environments on the performance of Deep Learning algorithms. We conducted experiments on a five node CPU commodity cluster, and present the results of our research in the form of two case studies on the major research fields of cosmology and remote sensing.

In the first case study, we address the problem of spectroscopic redshift estimation in astronomy, through a distributed perspective. We perform data distribution techniques in order to study the performance of a Convolutional Neural Network, considering both the number of training nodes and the way the data are distributed, while quantifying their effects via the metrics of training accuracy and training loss. In the second case study, we examine a research topic in the field of remote sensing. Our aim is to effectively split a multimodal Convolutional Neural Network used for multi-class land cover classification, that has a high number of parameters. Through model splits, we succeeded in effectively sharing the load of a Neural Network between the workers of our cluster and thus optimize CPU usage. We also managed to decrease the network traffic that happens due to frequent data transfers among the machines.