

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών
Υπολογιστών Διπλωματικές Εργασίες 2024
Καθηγητής Εμμανουήλ Βαρβαρίγος (vmanos@mail.ntua.gr)
Θέματα Διπλωματικών Εργασιών: 5G & Cloud Technologies

Θέμα 1: Ανίχνευση απόκλισης στην ακρίβεια μοντέλων μηχανικής μάθησης

Keywords: concept drift detection, machine learning, model accuracy, model performance

Οι εφαρμογές της μηχανικής μάθησης έχουν σημειώσει εκρηκτική άνοδο τα τελευταία χρόνια. Η σταθερή απόδοση των αλγορίθμων μηχανικής μάθησης είναι απαραίτητη προκειμένου να μπορούν τα μοντέλα αυτά να χρησιμοποιούνται με αξιόπιστο τρόπο σε κρίσιμες εφαρμογές. Για αυτό το λόγο, η ανίχνευση της απόκλισης στην ακρίβεια των αλγορίθμων είναι πολύ σημαντικό θέμα. Οι αποκλίσεις προκαλούνται συνήθως από σταδιακές αλλαγές στα δεδομένα εισόδου.

Σε αυτή τη διπλωματική θα μελετήσουμε αρχικά την έννοια της απόκλισης στην απόδοση ενός αλγορίθμου μηχανικής μάθησης και πως αυτή προκαλείται. Ακολούθως θα μελετήσουμε διάφορες τεχνικές ανίχνευσης της απόκλισης είτε στην απόδοση του αλγορίθμου, είτε της απόκλισης απευθείας στα δεδομένα εισόδου. Τέλος θα συγκρίνουμε τα αποτελέσματα των τεχνικών και πώς αυτά βοηθούν στην βελτίωση της σταθερότητας των μοντέλων μηχανικής μάθησης.

Επιθυμητές Γνώσεις: Python, Μηχανική μάθηση

Σχετική βιβλιογραφία

- S. Ackerman, P. Dube, E. Farchi, O. Raz, M. Zalmanovici, "Machine Learning Model Drift Detection Via Weak Data Slices," IEEE/ACM Third International Workshop on Deep Learning for Testing and Testing for Deep Learning (DeepTest), 2021
- L. Yang, A. Shami, "A lightweight concept drift detection and adaptation framework for IoT data streams," IEEE Internet of Things Magazine, 4(2), 96-101, 2021

Θέμα 2: Εντοπισμός βλαβών σε δικτυακές υποδομές με χρήση Μηχανικής Μάθησης

Keywords: network fault identification, network fault localization, machine learning, network tomography, complex networks

Η συνεχής εξέλιξη των δικτύων έχει ως αποτέλεσμα την αύξηση της πολυπλοκότητας διαχείρισής τους. Σε αυτό το περιβάλλον η έγκαιρη και αξιόπιστη ανίχνευση βλαβών είναι πολύ σημαντική για τη ορθή και αποδοτική λειτουργία του δικτύου. Οι ενσωματωμένοι μηχανισμοί ανίχνευσης και εντοπισμού βλαβών πολλές φορές είναι ανεπαρκής, και γι' αυτό απαιτούνται διατεματικές μετρικές.

Σε αυτή τη διπλωματική θα μελετηθεί το πρόβλημα της ανίχνευσης βλαβών κόμβων δικτύου χρησιμοποιώντας διατεματικές μετρικές μονοπατιών. Αρχικά θα μελετηθεί το υπόβαθρο των σύγχρονων δικτύων και θα οριστούν οι παράμετροι του προβλήματος. Ακολούθως, θα μελετηθούν τεχνικές Μηχανικής Μάθησης που μπορούν να χρησιμοποιηθούν για το συγκεκριμένο πρόβλημα. Τέλος θα γίνει εκτίμηση της ακρίβειας των τεχνικών για διαφορετικά και ρεαλιστικά σενάρια.

Επιθυμητές Γνώσεις: Δίκτυα επικοινωνιών, Python, Μηχανική μάθηση

Σχετική βιβλιογραφία

- S. M. Srinivasan, T. Truong-Huu, M. Gurusamy, "Machine learning-based link fault identification and localization in complex networks," IEEE Internet of Things Journal, 6(4), 6556-6566, 2019.

- N. Bartolini, T. He, V. Arrigoni, A. Massini, F. Trombetti, H. Khamfroush, "On fundamental bounds on failure identifiability by Boolean network tomography," IEEE/ACM Transactions on Networking, 28(2), 588-601, 2020.

Θέμα 3: Αλγόριθμοι αποτελεσματικής ανάθεσης cloud-native εφαρμογών σε ετερογενή και καταναμεμημένα περιβάλλοντα IoT-Edge-Cloud

Keywords: Hierarchical orchestration, cloud-edge resources, cloud native applications

Η αποτελεσματική διαχείριση των υπολογιστικών και δικτυακών πόρων θέτει προκλήσεις για την κατανομή των εφαρμογών στους κατάλληλους πόρους και τη διασφάλιση της παρεχόμενης ποιότητας υπηρεσιών.

Οι απαιτήσεις αυξάνονται με την εμφάνιση νέων εφαρμογών που περιλαμβάνουν δυναμικές ροές εργασίας που απαιτούν επεξεργασία σε πολλαπλά στάδια και ad-hoc υπολογισμούς. Το αντίστοιχο πρόβλημα γίνεται πολύπλοκο, όταν λαμβάνονται υπόψη οι περιορισμοί που θέτουν τα ετερογενή και δυναμικά καταναμεμημένα περιβάλλοντα πολλαπλών τεχνολογιών IoT, edge και cloud. Οι σημαντικότεροι περιορισμοί που πρέπει να λαμβάνονται υπόψη κατά τη διαδικασία βελτιστοποίησης είναι οι εκτιμήσεις σε χωρικές (απαιτήσεις επεξεργασίας) και χρονικές (εκτιμώμενος χρόνος επεξεργασίας) διαστάσεις, η ετερογένεια των πόρων, η δικαιοσύνη και η χρήση.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη αλγορίθμων που είναι σε θέση να αντιμετωπίσουν την εισαγόμενη αβεβαιότητα κατά τη δέσμευση των πόρων για την εξυπηρέτηση των διαφόρων εφαρμογών.

Επιθυμητές Γνώσεις: Δίκτυα επικοινωνιών, Python/Matlab, Μηχανική μάθηση.

Σχετική βιβλιογραφία

- R. McKenna, S. Herbein, A. Moody, T. Gamblin, and M. Taufer. "Machine learning predictions of runtime and IO traffic on high-end clusters". In: 2016 IEEE International Conference on Cluster Computing (CLUSTER). IEEE. 2016, pp. 255–258.

- M. R. Wyatt, S. Herbein, T. Gamblin, A. Moody, D. H. Ahn, and M. Taufer. "PRIONN: predicting runtime and IO using neural networks". In: Proceedings of the 47th International Conference on Parallel Processing. 2018, pp. 1–12.

- Bertsekas, D. "The auction algorithm: A distributed relaxation method for the assignment problem." Annals of operations research 14.1 (1988): 105-123.

- Y. Mao, Y. Fu, W. Zheng, L. Cheng, Q. Liu and D. Tao, "Speculative Container Scheduling for Deep Learning Applications in a Kubernetes Cluster," in IEEE Systems Journal, doi: 10.1109/JSYST.2021.3129974.

Θέμα 4: Ανάπτυξη αλγορίθμων για τη δέσμευση δικτυακών και υπολογιστικών πόρων για την εξυπηρέτηση των εφαρμογών βάση των προθέσεών τους

Keywords: intent based networking, cloud-edge resources, cloud native applications

Είναι γεγονός ότι οι νέες γενιές εφαρμογές γίνονται όλο και πιο πολύπλοκες με χαρακτηριστικά που διαφέρουν σημαντικά μεταξύ τους. Το πρόβλημα της εξυπηρέτησης των εφαρμογών γίνεται ακόμη πιο πολύπλοκο σε υποδομές που βασίζονται σε διαφορετικές τεχνολογίες. Η ανάπτυξη αλγορίθμων για την εξυπηρέτηση των εφαρμογών βάση των προθέσεών τους (intent based networking) κρίνεται απαραίτητη.

Στην παρούσα διπλωματική εργασία θα αναπτυχθούν αλγόριθμοι Μηχανικής Μάθησης για την εξυπηρέτηση εφαρμογών σε ετερογενή περιβάλλοντα βάση μιας γενικής περιγραφής των αναγκών τους (Intent Based)

Επιθυμητές Γνώσεις: Δίκτυα επικοινωνιών, Python/Matlab, Μηχανική μάθηση.

Σχετική βιβλιογραφία

- Jiachen Yang, Igor Borovikov, Hongyuan Zha, Hierarchical Cooperative Multi-Agent Reinforcement Learning with Skill Discovery

Θέμα 5: Χρήση τεχνικών βελτιστοποίησης για την επίλυση ερευνητικών προβλημάτων μηχανικής μάθησης

Keywords: Mixed Integer Linear Programming, ML model training.

Τα πολλαπλά χαρακτηριστικά που περιγράφουν τις δικτυακές και υπολογιστικές υποδομές εισάγουν μια επιπλέον πολυπλοκότητα για το σχεδιασμό μοντέλων μηχανικής μάθησης. Ο αριθμός τους αυξάνεται σημαντικά λαμβάνοντας υπόψη τα ετερογενή χαρακτηριστικά των cloud – edge υποδομών.

Κατά τη δημιουργία ενός μοντέλου Μηχανικής Μάθησης, η αφαίρεση άσχετων χαρακτηριστικών θα καταστήσει το μοντέλο ευκολότερο στην ερμηνεία και λιγότερο επιρρεπές στην υπερπροσαρμογή των δεδομένων.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη αλγορίθμων για την εύρεση του βέλτιστου υποσυνόλου χαρακτηριστικών, γνωστό ως πρόβλημα επιλογής καλύτερου υποσυνόλου.

Επιθυμητές Γνώσεις: Δίκτυα επικοινωνιών, Python/Matlab, Μηχανική μάθηση, Γραμμικός Ακέραιος Προγραμματισμός.

Σχετική βιβλιογραφία

- Dimitris Bertsimas, Angela King, Rahul Mazumder, Best Subset Selection via a Modern Optimization Lens

Θέμα 6: Δέσμευση πόρων για τη αποτελεσματική λειτουργία ασύρματων και οπτικών δικτύων (FiWi) για την παροχή υπηρεσιών από άκρο σε άκρο

Keywords: Cell planning, Fiber Wireless Networks, Network management, 5G Networks

Τα κινητά δίκτυα πέμπτης γενιάς (5G), στοχεύουν στην παροχή υπερ-υψηλής ταχύτητας μετάδοσης δεδομένων με παράλληλη μείωση έως και 90% της ενέργειας που καταναλώνεται εξασφαλίζοντας παράλληλα μικρή καθυστέρηση στη μεταφορά των δεδομένων. Για το λόγο αυτό, απαιτείται η σχεδίαση υποδομών υψηλής χωρητικότητας, που συνδυάζουν τις υπάρχουσες οπτικές και ασύρματες υποδομές. Βασικός στόχος είναι η παροχή διαφοροποιημένων υπηρεσιών, με παράλληλη αύξηση (έως 1000 φορές) της χωρητικότητας σε σύγκριση με τα 4G δίκτυα. Στα πλαίσια της διπλωματικής εργασίας θα αναπτυχθούν αλγόριθμοι σχεδίασης ετερογενών δικτύων που θα επιτρέπουν την πλήρη αξιοποίηση των υποδομών ασύρματων και οπτικών δικτύων, στοχεύοντας στην αύξηση της χρησιμοποίησης μιας τέτοιας υποδομής με παράλληλη μείωση της ενέργειας που καταναλώνεται. Η ανάπτυξη των αλγορίθμων θα γίνει με τη χρήση του προγραμματιστικού περιβάλλοντος MATLAB και τα εργαλεία που αυτό προσφέρει.

Σχετική βιβλιογραφία

- E. Björnson and L. Sanguinetti, "Scalable Cell-Free Massive MIMO Systems," in IEEE Transactions on Communications

- J. Zhang, Y. Xiao, D. Song, L. Bai and Y. Ji, "Joint Wavelength, Antenna, and Radio Resource Block Allocation for Massive MIMO Enabled Beamforming in a TWDM-PON Based Fronthaul

Θέμα 7: Αλγόριθμοι δέσμευσης πόρων για την εξυπηρέτηση cloud native εφαρμογών σε 5G υποδομές

Keywords: Cloud native, edge, Multiple-Access Edge Computing, 5G Networks, Virtual Network Embedding

Η ύπαρξη υπολογιστικών πόρων στα άκρα του δικτύου (edge computing) προσφέρει σημαντικά πλεονεκτήματα στην εξυπηρέτηση των αυστηρών και ετερογενών απαιτήσεων πολλών σύγχρονων εφαρμογών (IoT, Smart Factories, High-quality video, VR, etc). Οι edge υποδομές στοχεύουν να φέρουν πιο κοντά στους τελικούς χρήστες τα χαρακτηριστικά των cloud υποδομών με κύριο στόχο το συνδυασμό της αποτελεσματικής επεξεργασίας με χαμηλούς χρόνους απόκρισης. Όμως, οι διαθέσιμοι πόροι στις edge υποδομές χαρακτηρίζονται από μεγάλη ετερογένεια και περιορισμένες δυνατότητες σε σύγκριση με τις τυπικές cloud υποδομές. Παράλληλα, οι συγκεκριμένες υποδομές αποτελούν ένας από τους βασικότερους πυλώνες για την δημιουργία και υποστήριξη των σύγχρονων 5G υποδομών. Σε ένα τόσο ετερογενές και δυναμικό περιβάλλον, προκύπτει το πρόβλημα της απομόνωσης και μη αποδοτικής χρησιμοποίησης των ετερογενών edge πόρων. Στα πλαίσια της παρούσας διπλωματικής ως παραδείγματα εφαρμογών θα εξεταστούν σύγχρονες 5G εφαρμογές για την εξυπηρέτηση των οποίων απαιτείται η συντονισμένη ανάθεση υπολογιστικών πόρων τόσο για τις ίδιες τις εφαρμογές (MEC applications) όσο και για τις απαραίτητες 5G εικονικές δικτυακές λειτουργίες (5G Network Functions).

Οι εικονικές δικτυακές λειτουργίες (Network Function Virtualization - NFV) έχουν προταθεί και χρησιμοποιούνται ως ένας απαιτητικός μηχανισμός που επιτρέπει στους παρόχους δικτυακών υπηρεσιών να διαχειρίζονται και να παρέχουν κατ' απαίτηση δικτυακές λειτουργίες μέσω λογισμικού. Στόχος της εργασίας είναι η μελέτη μηχανισμών που θα υποστηρίζουν την βέλτιστη τοποθέτηση με δυναμικό και ευέλικτο τρόπο των επιμέρους δομικών τμημάτων (application functions) αυτών των εφαρμογών στους κατάλληλους edge πόρους λαμβάνοντας παράλληλα υπόψη τους περιορισμούς της υποδομής (δικτυακούς, επεξεργασίας και χώρου αποθήκευσης) και τις προδιαγραφές της εφαρμογής. Η ανάπτυξη των μηχανισμών θα στοχεύει στη βελτιστοποίηση της καθυστέρησης εξυπηρέτησης μέσω της κατάλληλης κατανομής των τμημάτων της εφαρμογής τόσο κάθετα (σε πόρους στα διαφορετικά επίπεδα των edge κόμβων) όσο και οριζόντια (σε διαφορετικούς κόμβους του ίδιου επιπέδου) στην υποδομή. Η ανάπτυξη των αλγορίθμων μπορεί να γίνει με χρήση του προγραμματιστικού περιβάλλοντος MATLAB ή Python και τα εργαλεία που αυτά προσφέρουν.

Σχετική βιβλιογραφία

- CNCF Cloud Native Definition: <https://github.com/cncf/toc/blob/main/DEFINITION.md>
- Qiang Duan, Intelligent and Autonomous Management in Cloud-Native Future Networks - A Survey on Related Standards from an Architectural Perspective, Future internet 2021.
- S. Shah, M. Gregory, S. Li, Cloud-Native Network Slicing Using Software Defined Networking Based Multi-Access Edge Computing: A Survey, IEEE Access, January 2021.

Θέμα 8: Αλγόριθμοι ανάθεσης εφαρμογών σε περιβάλλοντα που υποστηρίζουν προσεγγιστικούς υπολογισμούς μέσω υλικού

Keywords: approximate computing, resource allocation, edge-cloud resources

Οι επιταχυντές υλικού που υποστηρίζουν προσεγγιστικούς υπολογισμούς είναι ένα αναδυόμενο παράδειγμα για τον σχεδιασμό ενεργειακά αποδοτικών ή/και υψηλής απόδοσης υπολογιστικών υποδομών. Οι επιταχυντές υλικού κάνουν χρήση μιας πληθώρας υπολογιστικών τεχνικών που επιστρέφουν ένα πιθανώς ανακριβές αποτέλεσμα αντί για ένα εγγυημένα ακριβές αποτέλεσμα και που μπορούν να χρησιμοποιηθούν για εφαρμογές όπου ένα προσεγγιστικό αποτέλεσμα είναι επαρκές για το σκοπό του.

Ένα παράδειγμα τέτοιας κατάστασης είναι για μια μηχανή αναζήτησης όπου μπορεί να μην υπάρχει ακριβής απάντηση για ένα συγκεκριμένο ερώτημα αναζήτησης και ως εκ τούτου, πολλές απαντήσεις μπορεί να είναι αποδεκτές. Παρομοίως, η περιστασιακή αποβολή ορισμένων καρτέ σε μια εφαρμογή βίντεο μπορεί να περάσει απαρατήρητη λόγω των αντιληπτικών περιορισμών του ανθρώπου. Ο προσεγγιστικός υπολογισμός βασίζεται στην παρατήρηση ότι σε πολλά σενάρια, παρόλο που η εκτέλεση ακριβών υπολογισμών απαιτεί μεγάλη ποσότητα πόρων, το να επιτρέπεται περιορισμένη προσέγγιση μπορεί να προσφέρει δυσανάλογα κέρδη σε απόδοση και ενέργεια, ενώ παράλληλα επιτυγχάνεται αποδεκτή ακρίβεια αποτελεσμάτων.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη αλγορίθμων για την ανάθεση των εφαρμογών στους πόρους ενός ετερογενούς edge-cloud υπολογιστικού περιβάλλοντος που υποστηρίζει προσεγγιστικούς υπολογισμούς μέσω υλικού.

Σχετική βιβλιογραφία

- Ling Wang, Xiaohang Wang, Approximate communication for energy-efficient network-on-chip, Advances in Computers, Elsevier, Volume 124, 2022, Pages 151-215, ISSN 0065-2458, ISBN 9780323856881

Επιθυμητές Γνώσεις: Δίκτυα επικοινωνιών, Python/Matlab.