Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τομέας Επικοινωνιών, Ηλεκτρονικής και Συστημάτων Πληροφορικής

Ενιαία Κατανομή Πόρων και Επιλογή Δεδομένων με Βάση τη Σημαντικότητα σε Ασύρματα Δίκτυα NOMA Ομόσπονδης Μάθησης

Διπλωματική Εργασία

του

ΙΩΑΝΝΗ ΠΡΩΤΟΓΕΡΟΥ

Επιβλέπων: Συμεών Παπαβασιλείου

Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2024

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Επικοινωνιών, Ηλεκτρονικής και Συστημάτων Πληροφορικής

Ενιαία Κατανομή Πόρων και Επιλογή Δεδομένων με Βάση τη Σημαντικότητα σε Ασύρματα Δίκτυα NOMA Ομόσπονδης Μάθησης

Διπλωματική Εργασία

του

ΙΩΑΝΝΗ ΠΡΩΤΟΓΕΡΟΥ

Επιβλέπων: Συμεών Παπαβασιλείου
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 7η Οκτωβρίου 2024.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Συμεών Παπαβασιλείου	Ελένη Στάη	Βασίλειος Καρυώτης
Καθηγητής Ε.Μ.Π.	Επίκουρη Καθηγήτρια Ε.Μ.Π.	Επίκουρος Καθηγητής Ιόν. Παν.

Αθήνα, Οκτώβριος 2024

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Επικοινωνιών, Ηλεκτρονικής και Συστημάτων Πληροφορικής

(Υπογραφή)

.....

Ιωάννης Πρωτόγερος
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Περίληψη

Η Ομόπονδη Μάθηση αποτελεί έναν κατανεμημένο αλγόριθμο Μηχανικής Μάθησης κατά τον οποίο οι συμμετέχοντες εκπαιδεύουν συλλογικά ένα νευρωνικό δίκτυο χωρίς να διαμοιράζουν τα τοπικά τους δεδομένα, αλλά αποστέλλοντας μόνο τις ανανεώσεις του μοντέλου κατά την εκπαίδευσή του σε αυτά. Αυτό αποτελεί λύση στα προβλήματα ιδιωτικότητας και χρονικής καθυστέρησης που θα παρουσίαζε η αποστολή των δεδομένων σε έναν εξυπηρετητή για κεντροποιημένη εκπαίδευση, όμως εγείρονται ζητήματα για την ενεργειακή αποδοτικότητα, εφόσον οι συμμετέχοντες εκτελούν διεργασίες που περιέχουν αξιόλογο τόσο υπολογιστικό όσο και επικοινωνιακό φόρτο. Προς αυτήν την κατεύθυνση, καταφεύγουμε σε μεθόδους Κλασικής Βελτιστοποίησης για να διατυπώσουμε μεθοδολογίες και αλγορίθμους που επιλύουν αποδοτικά το πρόβλημα της ενεργειακά αποδοτικής ανάθεσης πόρων και επιλογής δεδομένων για την Ομόσπονδη Μάθηση, για τη μοντελοποίηση του οποίου θα αξιοποιήσουμε την μετρική της σημαντικότητας των δεδομένων που αποτελεί ένδειξη για την συνεισφορά του κάθε δείγματος στην μάθηση του μοντέλου.

Λέξεις Κλειδιά

Ομόσπονδη Μάθηση, Μη Ορθογωνική Πολλαπλή Πρόσβαση, Δίκτυα 6^{ης} γενιάς, Ενεργειακή Απόδοση, Κατανομή πόρων, Σημαντικότητα Δεδομένων, Μηχανική Μάθηση, Βελτιστοποίηση

Abstract

Federated Learning is a distributed Machine Learning framework in which participants collectively train a neural network without sharing their local data, by only sending updates of the model parameters during local training. This is a solution to the privacy and time delay problems of sending data to a server for centralized training. Still, it raises energy efficiency issues since participants perform processes with considerable computational and communication over-head. Towards this end, we resort to Classical Optimization techniques to formulate methodologies and algorithms that efficiently solve the problem of energy-efficient radio and computing resource allocation and data selection for Federated Learning, for the modeling of which we utilize the metric of data importance, which provides an indication of each sample's contribution to model learning.

Keywords

Federated Learning, Non-Orthogonal Multiple Access, 6^{th} generation Networks, Energy Efficiency, Resource Allocation, Data Importance, Machine Learning, Optimization

Στους συνοδοιπόρους μου

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στον καθηγητή κ. Συμεών Παπαβασιλείου για την επίβλεψη της παρούσας διπλωματικής εργασίας, αλλά και επειδή μου έδωσε την ευκαιρία να ανακαλύψω και να ασχοληθώ σε βάθος με ένα εξαιρετικά ενδιαφέρον θέμα. Θα ήθελα επιπλέον να ευχαριστήσω τους διδάκτορες Μάρω Διαμαντή και Δημήτρη Σπαθαράκη, για την βοήθεια και την καθοδήγηση που μου πρόσφεραν σε όλα τα στάδια της εκπόνησης της εργασίας. Η από κοινού μας προσπάθεια μου προσέφερε εφόδια που ευελπιστώ ότι θα φανούν χρήσιμα στην μετέπειτα πορεία μου ως Μηχανικός.

Αυτή η διπλωματική εργασία αντικατοπτρίζει το τελικό στάδιο της ακαδημαϊκής μου πορείας στην Σχολή Ηλεκτρολόγων Μηχανικών και Ηλεκτρονικών Υπολογιστών του Ε.Μ.Π., και γι'αυτό θα ήθελα να εκφράσω την βαθιά μου ευγνωμοσύνη στα άτομα που ήταν καθοριστικά στο να καταστεί αυτή επιτυχής και τελεσφόρα. Στην οικογένειά μου, για την στήριξη σε όλα τα χρόνια πριν και κατά τη διάρκεια των προπτυχιακών σπουδών μου. Στους φίλους και συμφοιτητές μου με τους οποίους δεθήκαμε και αλληλοστηριχθήκαμε κάτω από την κοινή μας εμπειρία και πρόκληση στις σπουδές μας. Τέλος, στον Θανάση, στον Γιώργο, και στη Νάσια, τα άτομα που παρά την απόσταση που μας χώριζε, κατέληξα να νιώθω πιο κοντά από ποτέ.

Ιωάννης Πρωτόγερος
Πειραιάς, Οκτώβριος 2024

Περιεχόμενα

Περίληψη 1

Abstract 3

Ευχαριστίες 7

Περιεχόμενα 10

Κατάλογος Σχημάτων 12

1 Εισαγωγή 13
1.1 Κίνητρο . . . 13
1.2 Διάρθρωση της Διπλωματικής Εργασίας . . . 14

2 Θεωρητικό Υπόβαθρο 17
2.1 Ομόσπονδη Μάθηση . . . 17
2.2 Ασύρματα Δίκτυα NOMA . . . 17
2.2.1 Το Πρόβλημα Ανάθεσης Ραδιοπόρων . . . 19
2.3 Επιλογή Δεδομένων με Βάση τη Σημαντικότητα . . . 19
2.4 Θεωρία Βελτιστοποίησης . . . 20
2.4.1 Κυρτή Βελτιστοποίηση . . . 20
2.4.2 Κλασματικός Προγραμματισμός . . . 23
2.4.3 Εναλλασσόμενη Βελτιστοποίηση . . . 24

3 Μοντελοποίηση Συστήματος 25
3.1 Σύστημα Ομόσπονδης Μάθησης . . . 25
3.1.1 Διαδικασία Ομόσπονδης Μάθησης . . . 25
3.2 Μοντέλο Επιλογής Δεδομένων . . . 27
3.2.1 Μοντέλο Συνελικτικού Νευρωνικού Δικτύου (CNN) . . . 28

3.2.2 Σημαντικότητα των Δεδομένων 29
3.3 Υπολογιστικό Μοντέλο 33
3.4 Τηλεπικοινωνιακό Μοντέλο 34
3.5 Μοντελοποίηση του Προβλήματος Βελτιστοποίησης 35

4 Επίλυση του Προβλήματος Βελτιστοποίησης 39
4.1 Διάσπαση του Προβλήματος σε Κυρτά Υποπροβλήματα 39
4.1.1 Μετατροπή του Κλασματικού Προβλήματος Βελτιστοποίησης με την τεχνική του Dinkelbach 39
4.1.2 Εναλλασσόμενη Βελτιστοποίηση 40
4.2 Επίλυση του Προβλήματος Ανάθεσης Δεδομένων 40
4.3 Επίλυση του Προβλήματος Ανάθεσης Υπολογιστικών Πόρων 43
4.4 Επίλυση του Προβλήματος Ανάθεσης Ραδιοπόρων 45
4.5 Επίλυση του Συνολικού Προβλήματος Βελτιστοποίησης 46

5 Διεξαγωγή Πειραμάτων 49
5.1 Παράμετροι του Συστήματος 49
5.2 Διεργασία Μηχανικής Μάθησης 50
5.2.1 Το Σύνολο Δεδομένων MNIST 50
5.2.2 Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου (CNN) 51
5.3 Τοπολογία Ασύρματου Δικτύου 51
5.4 Κατανομή των Δεδομένων Εκπαίδευσης στους Συμμεντέχοντες 53
5.5 Σημεία Αναφοράς για την Αξιολόγηση της Λύσης (Benchmarks) 53
5.5.1 Επίλυση του προβλήματος με Αλγόριθμο για Μη Κυρτή Βελτιστοποίηση 53
5.5.2 Εναλλακτική Στρατηγική Επιλογής Δεδομένων 54
5.5.3 Εναλλακτική Στρατηγική Ανάθεσης Ραδιοπόρων και Υπολογιστικών Πόρων 55

6 Αποτελέσματα 57
6.1 Σύγκλιση της Αναλυτικής Λύσης 57
6.2 Αξιολόγηση της Αναλυτικής Λύσης 63
6.3 Αξιολόγηση της Στρατηγικής Επιλογής Δεδομένων 66
6.4 Αξιολόγηση της Στρατηγικής Κοινής Ανάθεσης Ραδιοπόρων και Υπολογιστικών Πόρων 73

7 Συμπεράσματα και Μελλοντικές Επεκτάσεις 75

Βιβλιογραφία 77

Γλωσσάριο 83

Κατάλογος σχημάτων

2.1 Μια Τυπική Ροή Εργασίας σε Σύστημα Ομόσπονδης Μάθησης 18
2.2 Παράδειγμα που αναδεικνύει την ανάγκη χρήσης πολλαπλασιαστών Lagrange [20] 21

3.1 Αρχιτεκτονική συστήματος ομόσπονδης μάθησης με επιλογή δεδομένων 27
3.2 Ποιοτική γραφική αναπαράσταση της συνάρτησης ελάττωσης της τοπικής απώλειας g_n, με χαλαρωμένη σε συνεχή μεταβλητή B_n και παρεμβολή των τιμών της ως αυτών μιας τμηματικά γραμμικής συνάρτησης. 32

5.1 Εικόνες από το σύνολο MNIST [8] 51
5.2 Διάγραμμα της αρχιτεκτονικής του CNN που παράχθηκε μέσω της εφαρμογής Netron 52
5.3 Προσέγγιση της συνάρτησης σημαντικότητας $g_n(B_n)$ με λογαριθμική συνάρτηση για μια συσκευή. 54

6.1 Σύγκλιση της τιμής της αντικειμενικής συνάρτησης. Με κόκκινες τελείες είναι επισημειωμένες οι ανανεώσεις στην παράμετρο y του αλγορίθμου Dinkelbach. 58
6.2 Σύγκλιση της ενέργειας υπολογισμού 59
6.3 Σύγκλιση της ενέργειας για την μετάδοση 60
6.4 Σύγκλιση της ανάθεσης δεδομένων στους επεξεργαστές της συσκευής 2 60
6.5 Σύγκλιση στις συχνότητες λειτουργίας των επεξεργαστών της συσκευής 2 . . 61
6.6 Σύγκλιση στις ισχύς μετάδοσης των συσκευών 62
6.7 Ακρίβεια που επιτυγχάνεται μετά από 4 γύρους επικοινωνίας για αυξανόμενο αριθμό συσκευών στο δίκτυο ΟΜ 64
6.8 Μέση ενέργεια κατανάλωσης ανά γύρο επικοινωνίας για αυξανόμενο αριθμό συσκευών στο δίκτυο ΟΜ 64
6.9 Σύγκριση των χρόνων εκτέλεσης των διαφορετικών αλγορίθμων επίλυσης του προβλήματος βελτιστοποίησης 65

6.10 Χρόνος εκτέλεσης του προτεινόμενου αλγορίθμου κατά την κλιμάκωση του δικτύου ΟΜ 65
6.11 Αριθμός επιλεγμένων δειγμάτων (αριστερός y άξονας, μπλε χρώμα) και το ποσοστό του μέτρου της παραγώγου που συνιστούν ως επί το ολόκληρο σύνολο δεδομένων (δεξιός y άξονας, κόκκινο χρώμα) 66
6.12 Αθροιστική κατανομή της σημαντικότητας των δεδομένων σε μία συσκευή για τους 5 πρώτους γύρους επικοινωνίας του συστήματος ομόσπονδης μάθησης . . 67
6.13 Δεδομένα που ανατίθενται και συχνότητες που επιλέγονται για τους επεξεργαστές μίας συσκευής, κατά τη διάρκεια της ΟΜ 68
6.14 Ακρίβεια που επιτυγχάνεται για κάθε στρατηγική επιλογής δεδομένων, σε iid δεδομένα 69
6.15 Ακρίβεια που επιτυγχάνεται για κάθε στρατηγική επιλογής δεδομένων, σε non-iid δεδομένα 69
6.16 Ενεργειακή κατανάλωση για κάθε στρατηγική επιλογής δεδομένων (iid-δεδομένα) 70
6.17 Ενεργειακή κατανάλωση για κάθε στρατηγική επιλογής δεδομένων (non-iid δεδομένα) 71
6.18 Δεδομένα που επιλέγονται στο σύστημα για διαφορετικές τιμές του χρονικού ορίου T_{max} (Μέσος όρος 10 γύρων επικοινωνίας) 72
6.19 Ενέργεια ανά 1000 δείγματα που επιλέγονται στο σύστημα (μέσος όρος 10 γύρων επικοινωνίας) 72
6.20 Ενέργεια κατανάλωσης κάθε στρατηγικής ανάθεσης πόρων κατά τη διάρκεια της ΟΜ 74
6.21 Ενέργεια κατανάλωσης κατά την κλιμάκωση του δικτύου ΟΜ για διάφορες στρατηγικές ανάθεσης πόρων. Για κάθε διάταξη λαμβάνεται υπόψην ο μέσος όρος της ενέργειας στις 10 πρώτες εποχές. 74

1 Εισαγωγή

1.1 Κίνητρο

Η ενσωμάτωση της Τεχνητής Νοημοσύνης στον Κινητό Υπολογισμό και στα Ασύρματα Δίκτυα παρατηρείται ολοένα και πιο συχνά λόγω του αυξανόμενου αριθμού κινητών συσκευών και της διαθεσιμότητας τεράστιου όγκου δεδομένων. Ωστόσο, η κεντροποιημένη εκπαίδευση σε έναν τόσο μεγάλο πλήθος δεδομένων κατανεμημένων σε πολλές συσκευές θα προκαλούσε μεγάλη καθυστέρηση λόγω της μετάδοσης των δεδομένων σε έναν εξυπηρετητή, ενώ ταυτόχρονα εγείρεται το ζήτημα της ασφάλειας και της ιδιωτικότητας των δεδομένων.

Αυτό το πρόβλημα καλείται να αντιμετωπίσει η Ομόσπονδη Μάθηση - ΟΜ (Federated Learning - FL), που αποτελεί έναν κατανεμημένο αλγόριθμο μηχανικής μάθησης. Σε ένα δίκτυο ΟΜ, σε κάθε *γύρο επικοινωνίας* οι συμμετέχοντες εκπαιδεύουν όλοι την ίδια αρχιτεκτονική ενός νευρωνικού δικτύου, ο καθένας στο τοπικό του σύνολο δεδομένο, και έπειτα αποστέλλουν σε μία κεντρική οντότητα τις ανανεώσεις στις παραμέτρους του νευρωνικού δικτύου, προκειμένου να συναθροιστούν και να προκύψει ένα ανανεωμένο μοντέλο που έχει ουσιαστικά εκπαιδευτεί σε όλα τα δεδομένα στο σύστημα, χωρίς ωστόσο να υπάρξει καμία ανταλλαγή δεδομένων. Αυτή η διαδικασία επαναλαμβάνεται μέχρι τη σύγκλιση στις παραμέτρους του μοντέλου.

Παρά ωστόσο αυτό το πρόβλημα που επιλύει η Ομόσπονδη Μάθηση, ο υπολογισμός και η μετάδοση των ανανεώσεων στις παραμέτρους του μοντέλου αποτελεί κοστοβόρα διεργασία από άποψης χρόνου, αλλά και ενέργειας. Μάλιστα, σε συσκευές του Διαδικτύου των Πραγμάτων (Internet of Things - IoT) εγείρεται η ανάγκη επιστράτευσης πολλαπλών επεξεργαστών με δυνατότητα παράλληλης επεξεργασίας. Για τους παραπάνω λόγους, αποτελεί καίριο ερευνητικό αντικείμενο η ενεργειακά αποδοτική βελτιστοποίηση ενός ασύρματου δικτύου ΟΜ. Πολλές δουλειές προτείνουν αλγορίθμους ανάθεσης πόρων ή επιλογής δεδομένων προκειμένου να μειωθεί η καθυστέρηση, η ενεργειακή κατανάλωση, ή για να βελτιωθεί η απόδοση του μοντέλου.

Μελετώντας την επίδραση της υπολογιστικής διεργασίας (τοπική εκπαίδευση του νευρωνικού δικτύου) και της μετάδοσης στο ασύρματο δίκτυο στην συνολική καθυστέρηση και ενεργειακή κατανάλωση, γίνεται εμφανές ότι οι παράμετροι που αφορούν τον υπολογισμό (συχνότητες λειτουργίας επεξεργαστών) και την μετάδοση (ισχύς μετάδοσης συσκευών) αλληλεπιδρούν, εφόσον βρίσκονται υπό κοινούς χρονικούς και ενεργειακούς περιορισμούς. Δεν είναι

λοιπόν σπάνιο να βελτιστοποιούνται από κοινού αυτές οι παράμετροι σε οικοσυστήματα ομόσπονδης μάθησης. Ωστόσο, δεν έχει παρατηρηθεί κάποια σχετική μελέτη για την ταυτόχρονη βελτιστοποίηση των υπολογιστικών πόρων, των ισχύων μετάδοσης, και των δεδομένων. Αυτό ακριβώς θα είναι το αντικείμενο αυτής της διπλωματικής εργασίας, όπου σε αλγόριθμο ενιαίας ανάθεσης ραδιοπόρων και υπολογιστικών πόρων θα ενσωματωθεί η επιλογή δεδομένων με βάση τη σημαντικότητα, μία μετρική που εκφράζει την αλλαγή που μπορεί να φέρει ένα δείγμα στις παραμέτρους ενός μοντέλου, για να ελαττωθεί η ενεργειακή κατανάλωση ενός συστήματος Ομόσπονδης Μάθησης.

1.2 Διάρθρωση της Διπλωματικής Εργασίας

Η παρούσα διπλωματική εργασία περιγράφει την προτεινόμενη στρατηγική της ενεργειακά αποδοτικής ανάθεσης ραδιοπόρων και υπολογιστικών πόρων και επιλογής σημαντικών δεδομένων στις παρακάτω ενότητες.

- Ενότητα 2: Δίνεται το θεωρητικό υπόβαθρο πάνω στο οποίο βασίζεται αυτή η δουλειά, και παρουσιάζεται η χρήση του στη διεθνή βιβλιογραφία σε προβλήματα σχετικά με το αντικείμενο αυτής της διπλωματικής εργασίας. Γίνεται αναφορά στις απαραίτητες βασικές γνώσεις για την Ομόσπονδη Μάθηση, τα προβλήματα ανάθεσης πόρων στα τηλεπικοινωνιακά συστήματα NOMA, την επιλογή δεδομένων με βάση τη σημαντικότητα σε συστήματα Μηχανικής Μάθησης, και τη Θεωρία Βελτιστοποίησης (με έμφαση στους πολλαπλασιαστές Lagrange, στον αλγόριθμο του Dinkelbach, και στην εναλλασσόμενη βελτιστοποίηση).

- Ενότητα 3: Παρουσιάζεται αναλυτικά η μοντελοποίηση του συστήματος ομόσπονδης μάθησης. Αναλύεται η διαδικασία εκπαίδευσης στην ομόσπονδη μάθηση, και ορίζεται η έννοια της σημαντικότητας, η οποία χρησιμοποιείται για την επιλογή δεδομένων που θα ενσωματωθεί στην ΟΜ. Επίσης, καθορίζονται οι μετρικές για τον χρόνο και την ενέργεια που απαιτούνται για την ΟΜ, πάνω στις οποίες θα βασιστούμε για την διατύπωση του προβλήματος βελτιστοποίησης που θα κληθούμε να επιλύσουμε.

- Ενότητα 4: Το πρόβλημα βελτιστοποίησης που είναι διατυπωμένο στην προηγούμενη ενότητα υπόκειται σε μετατροπές και διαχωρίζεται σε υποπροβλήματα, προκειμένου να καταστεί εφικτή η αποδοτική επίλυσή του. Προτείνονται έπειτα αλγόριθμοι για την επίλυσή του που βασίζονται σε μεθόδους κλασικής βελτιστοποίησης.

- Ενότητα 5: Αναφέρονται τα χαρακτηριστικά του προσομοιωμένου συστήματος ομόσπονδης μάθησης και οι συνθήκες εκτέλεσης των πειραμάτων. Δίνονται οι τιμές των παραμέτρων του συστήματος, και περιγράφεται η διεργασία μηχανικής μάθησης πάνω στην οποία καλείται το σύστημα ΟΜ να εκπαιδευτεί.

- Ενότητα 6: Παρουσιάζονται τα αποτελέσματα των πειραμάτων. Ο προτεινόμενος αλγόριθμος αξιολογείται πολυσχιδώς και ξεχωριστά για κάθε συστατικό του στοιχείο. Δίνεται

έμφαση στις μετρικές της ενέργειας που καταναλώνεται τόσο σε έναν όσο και σε πολλαπλούς γύρους επικοινωνίας, και για κάποιες περιπτώσεις εξετάζεται η κλιμακωσιμότητα του συστήματος. Τα αποτελέσματα αντιπαραβάλλονται με άλλα πιθανά σενάρια κατά τα οποία οι αποφάσεις για τις ισχύς μετάδοσης, συχνότητες λειτουργίας, και για την επιλογή δεδομένων λαμβάνονται με διαφορετικούς τρόπους από τον προτεινόμενο που αποφασίζει και για τα τρία από κοινού, λαμβάνοντάς τα όλα υπόψην.

- Ενότητα 7: Συνοψίζεται το περιεχόμενο της διπλωματικής εργασίας και αναφέρονται πιθανές μελλοντικές προεκτάσεις για τη δουλειά που παρουσιάστηκε.

2

Θεωρητικό Υπόβαθρο

2.1 Ομόσπονδη Μάθηση

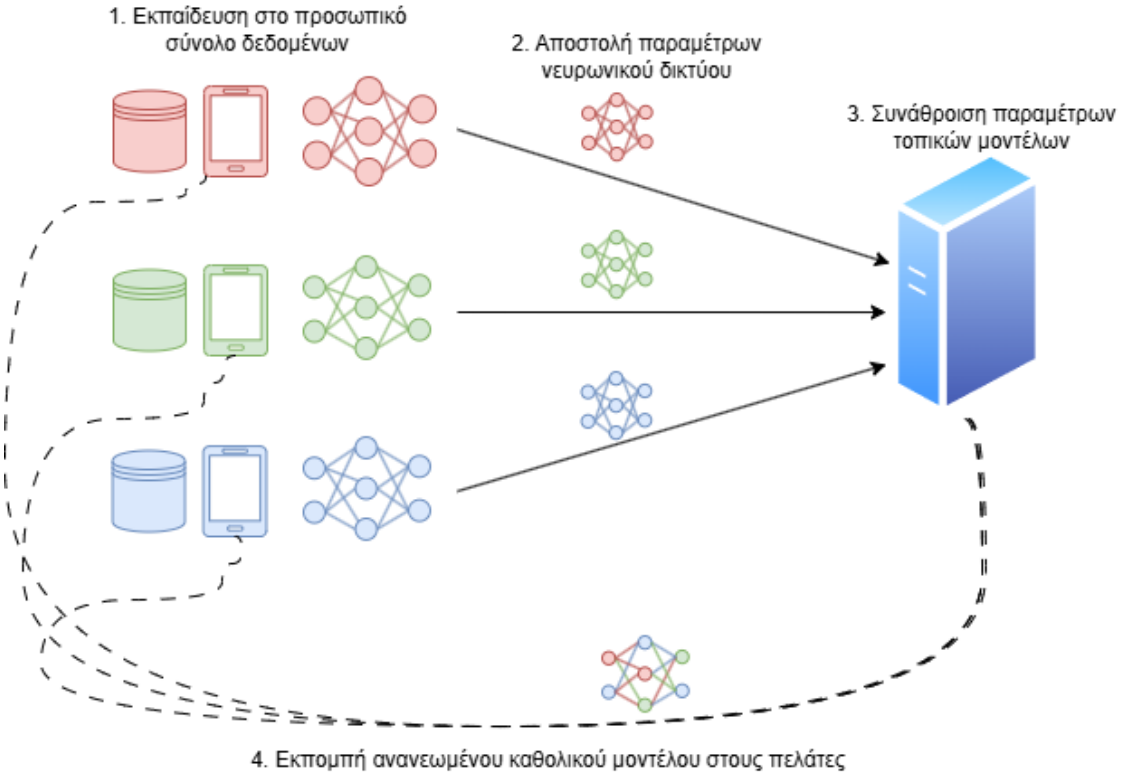
Η ραγδαία ανάπτυξη της Τεχνητής Νοημοσύνης τα τελευταία χρόνια έχει οδηγήσει σε αξιοσημείωτα επιτεύγματα σε ποικίλους τομείς, όπως η Όραση Υπολογιστών και η Επεξεργασία Φωνής και Φυσικής Γλώσσας [22]. Η επιτυχία της σε τέτοιους τομείς έγκειται στον τεράστιο όγκο δεδομένων που συλλέγεται για την εκπαίδευση. Ωστόσο, τα δεδομένα σε ασύρματα δίκτυα γενικά είναι κατανεμημένα σε μεγάλο αριθμό κινητών συσκευών. Για την πλήρη αξιοποίησή τους, η συμβατική προσέγγιση συνιστά την αποστολή τους σε έναν κεντρικό εξυπηρετητή στο Σύννεφο για την εκπαίδευση. Αυτό έχει δύο σημαντικά μειονεκτήματα: την έλλειψη ιδιωτικότητας και την μεγάλη καθυστέρηση στην επικοινωνία, που σταδιακά υποβαθμίζει την απόδοση της εκπαίδευσης.

Για την αντιμετώπιση των άνωθι προβλημάτων, προτείνεται η αποκεντρωμένη εκπαίδευση νευρωνικών δικτύων μέσω της Ομόσπονδης Μάθησης [30]. Η τυπική ροή εργασίας σε ένα δίκτυο ομόσπονδης μάθησης, που αναπαρίσταται και στο σχήμα 2.1, είναι η εξής: Οι συσκευές στα άκρα του δικτύου, αφού εκπαιδεύσουν ένα νευρωνικό δίκτυο στα τοπικά τους δεδομένα, αποστέλλουν στον διακομιστή μόνο τις επικαιροποιημένες παραμέτρους του (είτε βάρη είτε παραγώγους). Έπειτα, ο διακομιστής συναθροίζει τις παραμέτρους και εκπέμπει στους χρήστες το καινούριο μοντέλο ωστέ να εκπαιδευτεί ξανά κατά τον νέο γύρο επικοινωνίας.

Η ομόσπονδη μάθηση χρησιμοποιείται ευρέως σε τομείς που είναι απαραίτητη η διασφάλισης της ιδιωτικότητας των δεδομένων, όπως στον τομέα της Υγείας [32]. Ένα επίσης χαρακτηριστικό σενάριο χρήσης είναι η πρόβλεψη πληκτρολογούμενου κειμένου σε smartphones, όπου εταιρείες όπως η Google χρησιμοποιούν την ομόσπονδη μάθηση για να βελτιώσουν την πρόβλεψη κειμένου και την αυτόματη διόρθωση, χωρίς να θέσουν σε κίνδυνο τα προσωπικά δεδομένα των χρηστών [13].

2.2 Ασύρματα Δίκτυα NOMA

Η Μη Ορθογωνική Πολλαπλή Πρόσβαση (NOMA) είναι μια πρωτοποριακή τεχνική ραδιοπρόσβασης που έχει σχεδιαστεί για να βελτιώσει τις επιδόσεις των κυψελοειδών επικοινωνιών



Σχήμα 2.1: Μια Τυπική Ροή Εργασίας σε Σύστημα Ομόσπονδης Μάθησης

επόμενης γενιάς, ιδίως στα συστήματα 5G και Beyond-5G/B5G. Σε αντίθεση με τις παραδοσιακές μεθόδους Ορθογώνιας Πολλαπλής Πρόσβασης (Orthogonal Multiple Access- OMA), η τεχνική NOMA επιτρέπει σε πολλαπλούς χρήστες να μοιράζονται τους ίδιους πόρους συχνότητας διαφοροποιώντας τους μέσω επιπέδων ισχύος ή διαφορετικών κωδικών, βελτιώνοντας έτσι την αποδοτικότητα του φάσματος και τη συνδεσιμότητα [18, 45, 29].

Ένα κομβικό συστατικό της NOMA είναι η Διαδοχική Ακύρωση Παρεμβολών (Successive Interference Cancellation-SIC) στην πλευρά του παραλήπτη, που επιτρέπει πολλαπλούς χρήστες να χρησιμοποιήσουν ταυτόχρονα τους ίδιους φασματικούς πόρους μέσω της αποδοτικής διαχείρισης των παρεμβολών κατά την επεξεργασία του ληφθέντος σήματος.

Ο τρόπος με τον οποίον λειτουργεί η SIC είναι ο εξής [28, 16]: Ο δέκτης λαμβάνει ένα σήμα που αποτελεί την υπέρθεση όλων των σημάτων των χρηστών. Για να το αποκωδικοποιήσει, ξεκινάει με το σήμα του χρήστη που έχει το μεγαλύτερο κέρδος καναλιού, δηλαδή προσπαθεί να το εντοπίσει μέσα στο συνολικό σήμα που περιλαμβάνει τις παρεμβολές από όλες τις συσκευές. Όταν το εντοπίσει, το αφαιρεί από το συνολικό σήμα και συνεχίζει με το επόμενο. Με αυτόν τον τρόπο, το τελευταίο σήμα που θα αποκωδικοποιηθεί θα είναι εκείνο που έχει το μικρότερο κέρδος καναλιού. Δηλαδή, το ασθενέστερο σήμα θα αποκωδικοποιηθεί τελευταίο, χωρίς να υποστεί καμία παρεμβολή εφόσον όλα τα υπόλοιπα έχουν ήδη αφαιρεθεί.

2.2.1 Το Πρόβλημα Ανάθεσης Ραδιοπόρων

Η επικοινωνία ανάμεσα στις συσκευές και οι υπολογιστικές διεργασίες μπορεί να επιφέρει σημαντικό ενεργειακό κόστος και καθυστέρηση, λόγω των περιορισμένων πόρων επικοινωνίας και υπολογιστικών πόρων στις κινητές συσκευές. Η ανάθεση ραδιοπόρων αποτελεί σημαντικό πρόβλημα στον σχεδιασμό τηλεπικοινωνιακών συστημάτων. Ειδικά για την περίπτωση του power-domain NOMA, είναι καίριο το πρόβλημα της επιλογής των ισχύων μετάδοσης (γνωστό και ως power control) για την ικανοποίηση απαιτήσεων Ποιότητας Υπηρεσίας (Quality-of-Service - QoS) . Για την επίλυση τέτοιων προβλημάτων χρησιμοποιούνται τόσο κλασικές μέθοδοι βελτιστοποίησης [48], όσο και λύσεις βασισμένες στην Μηχανική Μάθηση και στην Θεωρία Παιγνίων [14, 40].

Είναι σύνηθες στη διατύπωση προβλημάτων ανάθεσης πόρων να βελτιστοποιούνται από κοινού οι υπολογιστικοί πόροι μαζί με τους ραδιοπόρους [49], λαμβάνοντας υπόψην τόσο τις υπολογιστικές διεργασίες όσο και το κομμάτι της επικοινωνίας, εφόσον αμφότερα συνεισφέρουν στο συνολικό ενεργειακό κόστος και στην χρονική καθυστέρηση και συνήθως συνδέονται μεταξύ τους μέσω των QoS απαιτήσεων. Αυτό είναι ιδιαίτερα καίριο για την ομόσπονδη μάθηση, εφόσον περιλαμβάνει μη αμελητέο φόρτο υπολογισμού και επικοινωνίας, και γι'αυτό ποικίλες δουλειές το εξετάζουν κατά την μελέτη της ενεργειακά αποδοτικής ομόσπονδης μάθησης. [47, 46].

2.3 Επιλογή Δεδομένων με Βάση τη Σημαντικότητα

Η δραματική αύξηση των διαθέσιμων δεδομένων εκπαίδευσης έχει καταστήσει εφικτή τη χρήση βαθέων νευρωνικών δικτύων, η οποία με τη σειρά της έχει αποφέρει σημαντικές εξελίξεις σε πολλούς τεχνολογικούς τομείς, όπως στην όραση υπολογιστών και στην επεξεργασία φυσικής γλώσσας. Ωστόσο, λόγω της πολυπλοκότητας του προβλήματος βελτιστοποίησης που προκύπτει, το υπολογιστικό κόστος αποτελεί βασικό ζήτημα στην εκπαίδευση μεγάλων αρχιτεκτονικών.

Κατά την εκπαίδευση τέτοιων μοντέλων, φαίνεται σχετικά εύκολα ότι δεν είναι όλα τα δείγματα εξίσου σημαντικά- πολλά από αυτά χειρίζονται σωστά μετά από λίγες μόνο εποχές εκπαίδευσης και αυτά θα μπορούσαν να αγνοηθούν σε εκείνο το σημείο χωρίς να επηρεάσουν το τελικό μοντέλο.

Για το σκοπό αυτό, προτείνεται στο [19] η τεχνική της επιλογής δεδομένων με βάση μια μετρική που ορίζεται ως η σημαντικότητα για κάθε δείγμα, που επιταχύνει την εκπαίδευση οποιουδήποτε νευρωνικού δικτύου εστιάζοντας τον υπολογισμό στα δείγματα που θα επιφέρουν τη μεγαλύτερη αλλαγή στις παραμέτρους του.

Η σημαντικότητα των δεδομένων έχει ενσωματωθεί σε εργασίες που εστιάζουν στην βελτιστοποίηση συστημάτων επικοινωνιών για διεργασίες μηχανικής μάθησης [27], ενώ στην πρόσφατη δουλειά των [15] προτείνεται η χρήση της σημαντικότητας για την βελτιστοποίηση ανάθεσης πόρων σε οικοσύστημα ομόσπονδης μάθησης για την ελαχιστοποίηση της διάρκειας ενός γύρου επικοινωνίας. Στο [33] προτείνεται επίσης ένας ενιαίος αλγόριθμος ανάθεσης υπολογι-

στικών πόρων και επιλογής σημαντικών δεδομένων προς την μείωση της καθυστέρησης και της ενέργειας σε ένα σύστημα ιεραρχικής ομόσπονδης μάθησης (hierarchical federated learning). Ωστόσο, δεν έχει εντοπιστεί η διατύπωση κάποιας στρατηγικής κοινής ανάθεσης ραδιοπόρων, υπολογιστικών πόρων, και επιλογής δεδομένων σε ασύρματα δίκτυα NOMA ομόσπονδης μάθησης.

Στο [15], η μετρική που χρησιμοποιείται για τον ορισμό της σημαντικότητας των δεδομένων είναι η εκτίμηση της παραγώγου $\partial \ell / \partial \boldsymbol{w}$ της συνάρτησης απώλειας ℓ που χρησιμοποιείται για την εκπαίδευση του νευρωνικού δικτύου με βάρη $\boldsymbol{w}$. Αυτή η εκτίμηση μπορεί να υπολογιστεί κατά την εμπρόσθια διάδοση, οπότε μπορεί να υπολογιστεί σε πρώτο στάδιο η σημαντικότητα για όλα τα δεδομένα, και μετά την επιλογή των σημαντικότερων δεδομένων, να συνεχίσει η διαδικασία της οπισθοδιάδοσης και της εκπαίδευσης μόνο για τα επιλεγμένα, "σημαντικότερα" δείγματα. Εφόσον η οπισθοδιάδοση έχει περίπου διπλάσιο υπολογιστικό φόρτο για τις περισσότερες αρχιτεκτονικές νευρωνικών δικτύων [19], αυτή η τεχνική μπορεί να ελαττώσει σημαντικά το χρονικό και το ενεργειακό κόστος για την εκπαίδευση μέσω της ομόσπονδης μάθησης.

2.4 Θεωρία Βελτιστοποίησης

Η Βελτιστοποίηση αποτελεί την διαδικασία απόκτησης του βέλτιστου αποτελέσματος κάτω από δεδομένες συνθήκες. Ο σχεδιασμός ενός συστήματος που βελτιστοποιεί ένα σύνολο μετρικών υπό ορισμένους περιορισμούς αποτελεί πρόβλημα που συναντάται ευρέως στην επιστήμη του Μηχανικού, και μπορεί να αφορά από φυσικά συστήματα όπως ένα μηχανοκίνητο όχημα έως μη φυσικά συστήματα όπως ένα που χρησιμοποιεί τεχνικές όρασης υπολογιστών για να αναγνωρίσει αν ένας όγκος είναι κακοήθης.

Επιθυμούμε αυτά τα συστήματα να είναι όσο πιο αποδοτικά γίνεται υπό τις υπάρχουσες συνθήκες, και με βάση αυτήν την ιδέα καλούμαστε να μοντελοποιήσουμε τις μετρικές και τους περιορισμούς που χαρακτηρίζουν το σύστημά μας με όση μεγαλύτερη ακρίβεια γίνεται. Έπειτα θα επιστρατεύσουμε υπάρχοντες αλγορίθμους βελτιστοποίησης για να βελτιώσουμε αυτές τις μετρικές, λαμβάνοντας υπόψην ότι επιθυμούμε αυτό να γίνει διατηρώντας μια εύλογη υπολογιστική πολυπλοκότητα.

2.4.1 Κυρτή Βελτιστοποίηση

Ένα κυρτό πρόβλημα βελτιστοποίησης έχει την παρακάτω μορφή [5]

$$\begin{aligned} &\underset{\boldsymbol{x}}{\text{minimize}} \quad f_0(\boldsymbol{x}) \\ &\text{subject to} \quad f_i(\boldsymbol{x}) \leq b_i, \;\; i = 1, \ldots, m \end{aligned} \tag{2.1}$$

όπου οι συναρτήσεις $f_0, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$ είναι κυρτές, δηλαδή ικανοποιούν τη σχέση

$$f_i(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) \leq \alpha f_i(\boldsymbol{x}) + \beta f_i(\boldsymbol{y}) \tag{2.2}$$

για όλα τα $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ και όλα τα $\alpha, \beta \in \mathbb{R}$ για τα οποία $\alpha + \beta = 1, \;\; \alpha \geq 0, \;\; \beta \geq 0$. Τα γνωστά προβλήματα των ελάχιστων τετραγώνων και του γραμμικού προγραμματισμού αποτελούν

ειδικές περιπτώσεις κυρτών προβλημάτων βελτιστοποίησης.

Ο λόγος που μας ενδιαφέρει η κυρτή βελτιστοποίηση είναι επειδή η μέθοδος των πολλαπλασιαστών Lagrange, που θα αναλυθεί αμέσως μετά, παρέχει ικανές και αναγκαίες συνθήκες για την έυρεση ολικά βέλτιστων λύσεων σε κυρτά προβλήματα.

Η Μέθοδος των Πολλαπλασιαστών Lagrange

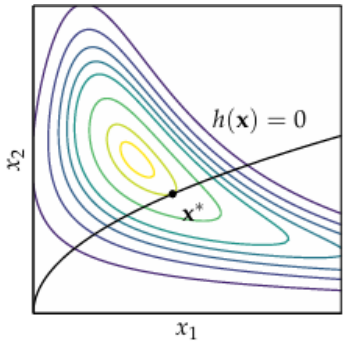
Τα παρακάτω ακολουθούν το κεφάλαιο 10.4 του βιβλίου [20] πάνω στους πολλαπλασιαστές Lagrange.

Η μέθοδος των πολλαπλασιαστών Lagrange χρησιμοποιείται για την βελτιστοποίηση μιας συνάρτησης με περιορισμούς ισότητας. Θα θεωρήσουμε πρόβλημα της μορφής

$$\begin{aligned} &\underset{\boldsymbol{x}}{\text{minimize}} \quad f(\boldsymbol{x}) \\ &\text{subject to} \quad h(\boldsymbol{x}) = 0 \end{aligned} \tag{2.3}$$

όπου οι f και h έχουν συνεχείς μερικές παραγώγους.

Μέσω των πολλαπλασιαστών Lagrange θέλουμε να υπολογίσουμε το σημείο όπου μία ισοσταθμική καμπύλη της f εφάπτεται με την καμπύλη της $h(\boldsymbol{x}) = 0$, όπως φαίνεται και στο παράδειγμα που δίνει το βιβλίο [20] (Σχήμα 2.2).



Σχήμα 2.2: Παράδειγμα που αναδεικνύει την ανάγκη χρήσης πολλαπλασιαστών Lagrange [20]

Εφόσον η κλίση μιας συνάρτησης σε κάποιο σημείο της είναι κάθετη στην ισοσταθμική καμπύλη που περνάει από αυτό το σημείο, γνωρίζουμε ότι η κλίση της h θα είναι κάθετη στην ισοσταθμική καμπύλη $h(x) = 0$. Επομένως, χρειάζεται να βρούμε σε ποιο σημείο η κλίση της f και η κλίση της h ευθυγραμμίζονται.

Επομένως, αναζητούμε το βέλτιστο σημείο $\boldsymbol{x}$ όπου ικανοποιείται ο περιορισμός

$$h(x) = 0 \tag{2.4}$$

και οι παράγωγοι βρίσκονται στην ίδια κατεύθυνση

$$\nabla f(\boldsymbol{x}) = \lambda \nabla h(\boldsymbol{x}) \tag{2.5}$$

για κάποιον Λαγκραντζιανό πολλαπλασιαστή , που είναι απαραίτητος επειδή οι κλίσεις μπορεί να έχουν διαφορετική κλίμακα.

Μπορούμε να διατυπώσουμε την Λαγκραντζιανή $\mathcal{L}$, που αποτελεί συνάρτηση ως προς τις μεταβλητές απόφασης του προβλήματος και του πολλαπλασιαστή

$$\mathcal{L}(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) - \lambda h(\boldsymbol{x}) \tag{2.6}$$

Επιλύοντας την $\nabla\mathcal{L} = 0$ προκύπτουν οι λύσεις των εξισώσεων 2.4 και 2.5. Οποιαδήποτε λύση μπορεί να αποτελεί τοπικό ελάχιστο, ή σημείο καμπής. Αν ωστόσο οι συναρτήσεις f και h είναι κυρτές, τότε μπορούμε να είμαστε βέβαιοι ότι αυτό το σημείο πρόκειται για ολικό ελάχιστο.

Τα παραπάνω επεκτείνονται και για πολλές ισότητες που αντιστοιχούν οι καθεμία σε έναν πολλαπλασιαστή Lagrange, αλλά και για προβλήματα με πολλαπλούς περιορισμούς-ανισότητες. Στην γενική περίπτωση, θεωρούμε το πρόβλημα

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{g}(\boldsymbol{x}) \leq 0 \\ & \boldsymbol{h}(\boldsymbol{x}) = 0 \end{aligned} \tag{2.7}$$

με κάθε στοιχείο του $\boldsymbol{g}$ να αποτελεί περιορισμό ανισότητας και κάθε στοιχείο του $\boldsymbol{h}$ να αποτελεί περιορισμό ισότητας. Τώρα λοιπόν εκτός από τους πολλαπλασιαστές Lagrange λ_i για τους περιορισμούς ισότητας $\boldsymbol{h}$ θα θεωρήσουμε και τους πολλαπλασιαστές μ_j για τους περιορισμούς ανισότητας $\boldsymbol{g}$.

Για τους περιορισμούς ανισότητας και τους αντίστοιχους πολλαπλασιαστές μ_j μπορούμε να θεωρήσουμε δύο περιπτώσεις: Σε μία βέλτιστη λύση είτε ο περιορισμός θα ισχύει με ισότητα (οπότε τον χαρακτηρίζουμε ως *ενεργό*), οπότε θα ισχύει η συνθήκη για την παράγωγο της αντικειμενικής συνάρτησης όπως και για τους περιορισμούς ισότητας (εξ. 2.5), είτε ο περιορισμός δεν θα ισχύει με ισότητα, οπότε το πρόβλημα θα είναι σαν να μην επιδέχεται τον σχετικό περιορισμό, και η συνθήκη Lagrange

$$\nabla f + \mu_j \nabla g_j = \mathbf{0} \tag{2.8}$$

θα ισχύει για $\mu_j = 0$.

Με βάση αυτό μπορούμε να ορίσουμε την γενικευμένη Λαγκραντζανιανή συνάρτηση, εφόσον αποδεικνύεται ότι για κάθε περιορισμό $g(\boldsymbol{x})$ μπορούμε να εισαγάγουμε μια γραμμική ποινή στην Λαγκραντζιανή συνάρτηση της μορφής $\mu g(\boldsymbol{x})$, που ωθεί την αντικειμενική συνάρτηση προς την εφικτότητα, αρκεί να ισχύει ότι $\mu > 0$.

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_i \mu_i g_i(\boldsymbol{x}) + \sum_j \lambda_j h_j(\boldsymbol{x}) \tag{2.9}$$

Η βελτιστοποίηση αυτού του προβλήματος περιλαμβάνει την εύρεση των κρίσιμων σημείων $\boldsymbol{x}^*$ για τα οποία ισχύουν οι παρακάτω συνθήκες, που είναι γνωστές ως οι *συνθήκες Karush-Kuhn-Tucker*.

- **Εφικτότητα**: Όλοι οι περιορισμοί ικανοποιούνται

$$\boldsymbol{g}(\boldsymbol{x}^*) \leq 0 \tag{2.10}$$

$$\boldsymbol{h}(\boldsymbol{x}^*) = 0 \tag{2.11}$$

- **Δυϊκή εφικτότητα**: Οι πολλαπλασιαστές μ_j για τους περιορισμούς ανισότητας "ωθούν" προς την πλευρά που ισχύει η ανισότητα

$$\boldsymbol{\mu} \geq 0 \tag{2.12}$$

- **Συμπληρωματική χαλαρότητα**: Οι πολλαπλασιαστές μ_j για τους περιορισμούς ανισότητας αναλαμβάνουν να αντικαταστήσουν τη χαλαρότητα στην συνάρτηση απώλειας Lagrange όταν αυτοί οι περιορισμοί ισχύουν χωρίς την ισότητα. Θα πρέπει λοιπόν για κάθε περιορισμό ανισότητας είτε αυτός να ισχύει με ισότητα είτε ο πολλαπλασιαστής να είναι μηδενικός. (το $\odot$ συμβολίζει την πράξη πολλαπλασιασμού για κάθε στοιχείο δύο διανυσμάτων)

$$\boldsymbol{\mu} \odot \boldsymbol{g} = \mathbf{0} \tag{2.13}$$

- **Στασιμότητα**: Οι ισοσταθμικές καμπύλες της αντικειμενικής συνάρτησης εφάπτονται σε κάθε ενεργό περιορισμό.

$$\nabla \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \nabla f(\boldsymbol{x}^*) + \sum_i \mu_i \nabla g_i(\boldsymbol{x}^*) + \sum_j \lambda_j \nabla h_j(\boldsymbol{x}^*) = 0 \tag{2.14}$$

2.4.2 Κλασματικός Προγραμματισμός

Ο Κλασματικός Προγραμματισμός (Fractional Programming) αναφέρεται στην οικογένεια προβλημάτων βελτιστοποίησης που περιέχουν όρους πηλίκου. Η ύπαρξή του μπορεί να εντοπιστεί ακόμα και σε μια δημοσίευση του von Neumann το 1937 για την οικονομική ανάπτυξη [31]. Έκτοτε έχει μελετηθεί εκτενώς η χρήση του σε τομείς της οικονομίας, της θεωρίας πληροφορίας, της οπτικής, της θεωρίας γραφημάτων και της επιστήμης των υπολογιστών [37, 41, 1].

Ο Αλγόριθμος Dinkelbach

Αυτή η κλασική τεχνική, που πρώτα προτάθηκε στο [10], επαναδιατυπώνει το πρόβλημα που αποτελείται από ένα πηλίκο

$$\begin{aligned} &\underset{\boldsymbol{x}}{\text{maximize}} \quad \frac{A(\boldsymbol{x})}{B(\boldsymbol{x})} \\ &\text{subject to} \quad \boldsymbol{x} \in \mathcal{X} \end{aligned} \tag{2.15}$$

στο εξής πρόβλημα:

$$\begin{aligned} &\underset{\boldsymbol{x}}{\text{maximize}} \quad A(\boldsymbol{x}) - yB(\boldsymbol{x}) \\ &\text{subject to} \quad \boldsymbol{x} \in \mathcal{X} \end{aligned} \tag{2.16}$$

χρησιμοποιώντας την βοηθητική μεταβλητή y, που ανανεώνεται σύμφωνα με τον τύπο

$$y[t+1] = \frac{A(\boldsymbol{x}[t])}{B(\boldsymbol{x}[t])} \tag{2.17}$$

Σημειώνεται ότι το διάνυσμα $\boldsymbol{x}$ αποτελεί το σύνολο μεταβλητών και το $\mathcal{X}$ αποτελεί το εφικτό σύνολο λύσεων (feasible set). Όταν το πρόβλημα 2.15 είναι concave-convex, δηλαδή ο αριθμητής $A(\boldsymbol{x})$ είναι κοίλος και ο παρανομαστής $B(\boldsymbol{x})$ είναι κυρτός, τότε το πρόβλημα βελτιστοποίησης 2.16 είναι κυρτό πρόβλημα, οπότε μπορεί να επιλυθεί πολύ πιο εύκολα σε σχέση με το αρχικό. Η μετατροπή του Dinkelbach έχει το πλεονέκτημα συγκριτικά με άλλες τεχνικές όπως η μετατροπή Charnes-Cooper ότι δεν εισάγει επιπρόσθετους περιορισμούς [39].

2.4.3 Εναλλασσόμενη Βελτιστοποίηση

Έστω $f : \mathbb{R}^s \to \mathbb{R}$ ένα βαθμωτό πεδίο του $\mathbb{R}$, και έστω ότι διαμερίζουμε το σύνολο των μεταβλητών $\boldsymbol{x} = \{x_1, \ldots, x_s\}$ σε t μη επικαλυπτόμενα υποσύνολα ως εξής: $\boldsymbol{x} = \{X_1, \ldots, X_t\}$, με $X_i \in \mathbb{R}^{k_i}, i = 1, \ldots, t$, και $\sum_{i=1}^{t} k_i = s$. *Εναλλασσόμενη βελτιστοποίηση* (Alternating optimization - AO) ονομάζεται η επαναληπτική διαδικασία ελαχιστοποίησης (ή μεγιστοποίησης) της συνάρτησης $f(X_1, X_2, \ldots, X_t)$ ως προς όλες τις μεταβλητές από κοινού, μέσω της εναλλαγής ανάμεσα στην βελτιστοποίηση της συνάρτησης ως προς κάθε ξεχωριστό υποσύνολο μεταβλητών [4]. Ο αλγόριθμος AO αποτελεί την βάση των αλγορίθμων ομαδοποίησης k-μέσων (k-means clustering) [3] και του Expectation-Maximization (EM) για την εκτίμηση των παραμέτρων σε μία μίξη Γκαουσιανών κατανομών [44].

Τυπικά, αυτή η μέθοδος ενδείκνυται για χρήση όταν στην δομή του προβλήματος παρατηρείται ένας φυσικός διαχωρισμός των μεταβλητών. Εν γένει, αυτή η μέθοδος βοηθάει στην διάσπαση ενός πολύπλοκου προβλήματος σε πολλαπλά, απλούστερα προβλήματα, προκειμένου να καταλήξουμε με την επαναληπτική επίλυσή τους σε μια υποβέλτιστη λύση, κάτι που όμως γίνεται με πολύ πιο γρήγορη σύγκλιση, ειδικά αν τα υποπροβλήματα είναι αισθητά πιο εύκολα και έχουν κλειστού τύπου λύσεις [25].

3

Μοντελοποίηση Συστήματος

3.1 Σύστημα Ομόσπονδης Μάθησης

Θα θεωρήσουμε ένα σύστημα που αποτελείται από σύνολο χρηστών/συσκευών $\mathcal{N}$, όπου

$$\mathcal{N} = \{1, 2, \ldots, N\}$$

Η κάθε συσκευή n διαθέτει ένα προσωπικό σύνολο δεδομένων με πληθάριθμο D_n. Ορίζουμε

$$\mathcal{D}_n = \{\boldsymbol{x}_i, y_i\}_{i=1}^{D_n}$$

το προσωπικό σύνολο δεδομένων του χρήστη n, όπου $\boldsymbol{x}_i, y_i$ είναι τα δείγματα και οι αντίστοιχες κλάσεις που απαρτίζουν αυτό το σύνολο.

Φυσικά, ισχύει για το σύνολο $\mathcal{D}$ (με $|\mathcal{D}| = D$) που συνιστά την συλλογή όλων των δεδομένων στο σύστημα ομόσπονδης μάθησης ότι

$$\mathcal{D} = \bigcup_{i=1}^{n} \mathcal{D}_n = \{\boldsymbol{x}_i, y_i\}_{i=1}^{D}$$

3.1.1 Διαδικασία Ομόσπονδης Μάθησης

Σε κάθε γύρο επικοινωνίας στη διαδικασία της ομόσπονδης μάθησης η κάθε συσκευή θα εκπαιδεύει τοπικά ένα νευρωνικό δίκτυο πάνω σε ένα **υποσύνολο** του προσωπικού συνόλου δεδομένων $\mathcal{B}_n \subseteq \mathcal{D}_n$. Αυτό αποσκοπεί στη μείωση της απαιτούμενης ενέργειας ανά γύρο σε σύγκριση με την εκπαίδευση σε ολόκληρο το σύνολο $\mathcal{D}_n$. Για να επιτευχθεί όμως αυτός ο στόχος και στη συνολική διαδικασία της ομόσπονδης μάθησης, θα πρέπει η διαδικασία της επιλογής δεδομένων να είναι τέτοια ώστε να επιταχύνει την σύγκλιση της εκπαίδευσης. Ο τρόπος με τον οποίο θα καθοριστεί αυτό το υποσύνολο $\mathcal{B}_n$ θα αναλυθεί στην επόμενη υποενότητα.

Μετά την επιλογή του υποσυνόλου $\mathcal{B}_n$ για τη συσκευή n, ακολουθεί η εκπαίδευση του τοπικού μοντέλου για έναν αριθμό εποχών, που βασίζεται στην ανανέωση των βαρών του νευρωνικού δικτύου σύμφωνα με τον τύπο

$$\boldsymbol{w}_n[k+1] = \boldsymbol{w}_n[k] - \eta \nabla L_n(\boldsymbol{w}_n[k], \mathcal{B}_n), \tag{3.1}$$

όπου η είναι ο ρυθμός μάθησης ,και L_n η τοπική συνάρτηση απώλειας.

Συνάθροιση τοπικών μοντέλων

Για την συνάθροιση των τοπικών μοντέλων που προκύπτουν μετά την εκπαίδευση κάθε συσκευής στο προσωπικό της σύνολο δεδομένων, θα χρησιμοποιηθεί ο αλγόριθμος FedAvg, ή μια παραλλαγή του λόγω της διαδικασίας επιλογής των δεδομένων εκπαίδευσης σε κάθε γύρο επικοινωνίας της ομόσπονδης μάθησης. Σύμφωνα με αυτόν τον αλγόριθμο, τα βάρη των τοπικών νευρωνικών δικτύων $\boldsymbol{w}_n$ μεταφέρονται μετά την τοπική εκπαίδευση στον κεντρικό διακομιστή, όπου γίνεται η συνάθροιση των μοντέλων με βάση τον τύπο

$$\boldsymbol{w} = \frac{1}{\sum_{n \in \mathcal{N}} |\mathcal{B}_n|} \sum_{n \in \mathcal{N}} |\mathcal{B}_n| \cdot \boldsymbol{w}_n \tag{3.2}$$

Επειδή τα νευρωνικά δίκτυα θεωρούμε ότι αρχικοποιούνται στις ίδιες παραμέτρους για όλες τις συσκευές, το να ανανεώνονται τα βάρη του συνολικού μοντέλου σύμφωνα με τον τύπο 3.2 είναι ισοδύναμο με το να συναθροίζονται οι τοπικές παράγωγοι $\boldsymbol{g}_n = \nabla L_n(\boldsymbol{w}_n, \mathcal{D}_n)$ σύμφωνα με τον τύπο

$$\boldsymbol{g} = \frac{1}{\sum_{n \in \mathcal{N}} |\mathcal{B}_n|} \sum_{n \in \mathcal{N}} |\mathcal{B}_n| \cdot \boldsymbol{g}_n \tag{3.3}$$

και έπειτα η ανανέωση των παραμέτρων του νευρωνικού δικτύου να γίνεται σύμφωνα με τον τύπο

$$\boldsymbol{w}[k+1] = \boldsymbol{w}[k] - \eta \boldsymbol{g} \tag{3.4}$$

όπου k είναι ο αριθμός της συνολικής επανάληψης.

Συνολικά, η λειτουργία του συστήματος ομόσπονδης μάθησης με επιλογή δεδομένων με βάση τη σημαντικότητα περιγράφεται με τα εξής βήματα, όπως απεικονίζονται και στο σχήμα 3.1:

- ① Forward propagation σε κάθε συσκευή για το προσωπικό σύνολο δεδομένων της $\mathcal{D}_n$ και υπολογισμός της σημαντικότητας κάθε δείγματος, με τον τρόπο που θα περιγραφεί στην υποενότητα 3.2.

- ② Ασύρματη μετάδοση από κάθε συσκευή στον διακομιστή της μετρικής σημαντικότητας $\sigma_{n,m}$ για κάθε δεδομένο. Επειδή αυτή η μετρική αποτελεί βαθμωτό, θεωρούμε ότι τα δεδομένα μου αποστέλλονται έχουν αρκετά μικρό μέγεθος για να θεωρηθεί αμελητέλος ο χρόνος και η ενέργεια μετάδοσης σε αυτό το βήμα.

- ③ Επιλογή δεδομένων και ανάθεση πόρων - Ο διακομιστής με βάση την κατάσταση του καναλιού, τα χαρακτηριστικά των επεξεργαστών των συσκευών, και την σημαντικότητα

των δεδομένων επιλέγει και στέλνει πίσω το πλήθος των δεδομένων B_n που θα ανατεθούν σε κάθε συσκευή, καθώς και τις συχνότητες επεξεργασίας f_n^q και ισχύς μετάδοσης p_n.

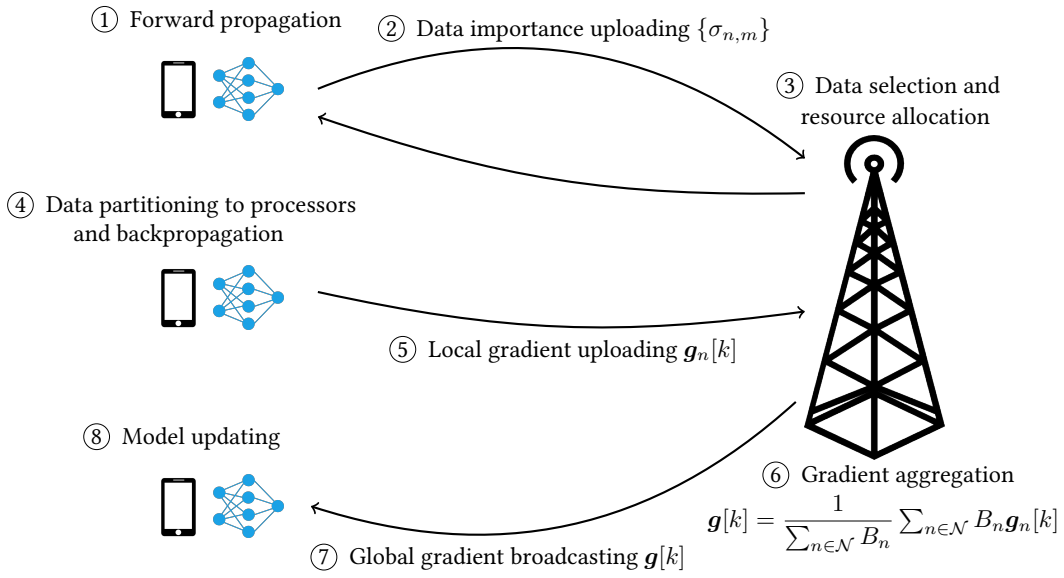
- ④ Κατακερματισμός των επιλεγμένων δεδομένων $\mathcal{B}_n$ στους Q_n επεξεργαστές σε κάθε συσκευή και εκτέλεση του backpropagation σε αυτό το σύνολο δεδομένων προς τον υπολογισμό των τοπικών παραγώγων $\boldsymbol{g}_n$. Αυτό μπορεί να επαναληφθεί για συγκεκριμένο αριθμό τοπικών εποχών.

- ⑤ Μετάδοση των υπολογισμένων παραγώγων μέσω του ασύρματου δικτύου στον διακομιστή.

- ⑥ Συνάθροιση των τοπικών παραγώγων προς τον υπολογισμό της συνολικής παραγώγου $\boldsymbol{g}$. Αυτό συνιστά μια απλή πράξη υπολογισμού μέσου όρου, οπότε θεωρούμε αυτήν τη διεργασία αμελητέα.

- ⑦ Εκπομπή της παραγώγου $\boldsymbol{g}$ σε όλες τις συσκευές. Η χρονική καθυστέρηση και η απαιτούμενη ενέργεια για αυτήν την εργασία δεν επηρεάζεται από τη διαδικασία βελτιστοποίησης που θα διατυπώσουμε, οπότε δεν θα συμπεριληφθεί στην ανάλυση.

- ⑧ Ανανέωση των παραμέτρων του καθολικού νευρωνικού δικτύου σύμφωνα με τον τύπο 3.4



Σχήμα 3.1: Αρχιτεκτονική συστήματος ομόσπονδης μάθησης με επιλογή δεδομένων

3.2 Μοντέλο Επιλογής Δεδομένων

Η διαδικασία της ομόσπονδης μάθησης περιλαμβάνει σε κάθε γύρο επικοινωνίας την εκπαίδευση ενός νευρωνικού δικτύου σύμφωνα με τον αλγόριθμο οπισθοδιάδοσης (backpropagation).

Αυτό απαιτεί τον υπολογισμό των παραγώγων από κάθε συσκευή πάνω στα τοπικά δεδομένα, το οποίο λόγω των περιορισμένων υπολογιστικών πόρων αποδεικνύεται ιδιαίτερα χρονοβόρο και κοστοβόρο από ενεργειακής άποψης. Προς αυτήν την κατεύθυνση, θα βασιστούμε στην δουλειά των [15] για να ενσωματώσουμε τη στρατηγική επιλογής των πιο σημαντικών δεδομένων για την τοπική εκπαίδευση στην συνολική διαδικασία ελαχιστοποίησης της ενέργειας κατά την ομόσπονδη μάθηση.

3.2.1 Μοντέλο Συνελικτικού Νευρωνικού Δικτύου (CNN)

Σε αυτήν την εργασία θα ασχοληθούμε με την χρήση συνελικτικών νευρωνικών δικτύων (CNNs), ωστόσο η γενική ιδέα μπορεί να επεκταθεί σε οποιοδήποτε μοντέλο μηχανικής μάθησης χρησιμοποιεί αλγόριθμο βασισμένο στο Gradient Descent για την εκπαίδευση.

Θα χρησιμοποιήσουμε τον συμβολισμό $\Psi(\boldsymbol{x}, \boldsymbol{w})$ για την έξοδο του CNN μοντέλου με διάνυσμα βαρών w. Το σφάλμα εκπαίδευσης του δείγματος $(\boldsymbol{x}_i, y_i)$ υπολογίζεται με την συνάρτηση απώλειας $\ell(\Psi(\boldsymbol{x_i}, \boldsymbol{w}), y_i)$. Έτσι, η τοπική συνάρτηση απώλειας για κάθε συσκευή δίνεται από την

$$L_n(\boldsymbol{w}, \mathcal{D}_n) = \sum_{(\boldsymbol{x}_i, y_i) \in \mathcal{D}_n} \ell(\Psi(\boldsymbol{x_i}, \boldsymbol{w}), y_i),\ \forall n \in \mathcal{N} \tag{3.5}$$

και η συνάρτηση της συνολικής απώλειας μπορεί να δοθεί από τον μέσο όρο όλων των τοπικών συναρτήσεων απώλειας ως εξής

$$L(\boldsymbol{w}) = \frac{1}{|\mathcal{D}|} \sum_{n \in \mathcal{N}} L_n(\boldsymbol{w}, \mathcal{D}_n) \tag{3.6}$$

Προς την ελαχιστοποίηση της συνάρτησης συνολικής απώλειας αποσκοπεί η εκπαίδευση των τοπικών μοντέλων, που με τη σειρά της στοχεύει στην ελαχιστοποίηση της τοπικής συνάρτησης απώλειας. Για αυτό χρησιμοποιείται ευρέως ο αλγόριθμος κατάβασης κλίσης, που περιλαμβάνει τον υπολογισμό της παραγώγου

$$\boldsymbol{g}_n = \nabla L_n(\boldsymbol{w}_n, \mathcal{D}_n) = \sum_{(\boldsymbol{x}_i, y_i) \in \mathcal{D}_n} \frac{\partial \ell(\Psi(\boldsymbol{x_i}, \boldsymbol{w}_n), y_i)}{\partial \boldsymbol{w}_n},\ \forall n \in \mathcal{N} \tag{3.7}$$

η οποία γίνεται σε δύο βήματα [11], με την εμπρόσθια διάδοση (forward pass), όπου γίνεται ο υπολογισμός του σφάλματος για κάθε δείγμα, και την οπισθοδιάδοση (backpropagation), κατά την οποία υπολογίζεται η παράγωγος με βάση την τιμή του σφάλματος. Έπειτα, η ανανέωση των παραμέτρων του νευρωνικού δικτύου γίνεται σύμφωνα με τον τύπο

$$\boldsymbol{w}[k+1] = \boldsymbol{w}[k] - \eta[k]\boldsymbol{g}[k],\ \forall n \in \mathcal{N}, \tag{3.8}$$

όπου k είναι η επανάληψη της τοπικής εκπαίδευσης, $\eta[k]$ είναι ο ρυθμός μάθησης κατά την $k-$οστή επανάληψη, και $\boldsymbol{g}[k]$ η συναθροισμένη, συνολική παράγωγος. Επειδή θεωρούμε ότι όλες οι συσκευές ξεκινούν

3.2.2 Σημαντικότητα των Δεδομένων

Όπως αναφέρεται και στο [51], η σημαντικότητα των δεδομένων μπορεί να συμπεριληφθεί στην διαδικασία της ανάθεσης πόρων προς την επιτάχυνση της διαδικασίας μάθησης. Σε αυτήν την εργασία θα χρησιμοποιήσουμε την επιλογή δεδομένων ως μέρος της στρατηγικής ελαχιστοποίησης της ενέργειας για τον υπολογιστικό και επικοινωνιακό φόρτο της διαδικασίας ομόσπονδης μάθησης.

Θα ορίσουμε την σημαντικότητα για τα δεδομένα βασιζόμενοι στον τρόπο ανανέωσης των παραμέτρων των μοντέλων, που επηρεάζει την επίδοση στην εκπαίδευση. Από την εξ. 3.8 είναι φανερό ότι το διάνυσμα των παραγώγων καθορίζει την ανανέωση των βαρών, υποδεικνύωντας ότι η παράγωγος μπορεί να αποτελέσει ένδειξη για την "σημαντικότητα" ενός δείγματος. Με μία μεγαλύτερη παράγωγο, το δεδομένο συνεισφέρει περισσότερο στην αλλαγή των παραμέτρων, και άρα είναι περισσότερο σημαντικό για τη σύγκλιση των βαρών του μοντέλου.

Ωστόσο, η πραγματική τιμή της παραγώγου προκύπτει μετά από τα βήματα του forward-pass και του backpropagation, οδηγώντας σε υπέρογκο υπολογιστικό φόρτο αν επιλέγαμε τα δεδομένα με βάση την πραγματική τιμή της παραγώγου. Είναι εφικτό όμως να προσεγγίσουμε το μέτρο του διανύσματος της παραγώγου χρησιμοποιώντας την απώλεια κάθε δείγματος μετά την εμπρόσθια διάδοση ως εξής, σύμφωνα με το [19]:

$$\|\boldsymbol{g}[k](\boldsymbol{x}_i, y_i)\|_2 = \left\| \frac{\partial \ell((\boldsymbol{x}_i, \boldsymbol{w}), y_i)}{\partial \boldsymbol{w}} \right\|_2 \approx \rho \left\| \frac{\partial \ell((\boldsymbol{x}_i, \boldsymbol{w}), y_i)}{\partial \boldsymbol{x}_i^L} \right\|_2, \tag{3.9}$$

όπου το $\boldsymbol{x}_i^L$ είναι η είσοδος στην συνάρτηση ενεργοποίησης του επιπέδου εξόδου του CNN, το ρ αποτελεί παράμετρο που εξαρτάται από το νευρωνικό δίκτυο, και η $\|\cdot\|$ είναι η L2 νόρμα. Στα [19, 17] αναφέρεται ότι η οπισθοδιάδοση στις περισσότερες αρχιτεκτονικές βαθέων νευρωνικών δικτύων απαιτεί περίπου διπλάσιες πράξεις κινητής υποδιαστολής (FLOPs) από την εμπρόσθια διάδοση. Εφόσον με αυτόν τον τρόπο προσέγγισης της τιμής της παραγώγου ο υπολογισμός κατά το backpropagation σταματάει στο επίπεδο εξόδου, μειώνεται αισθητά το υπολογιστικό κόστος για την εκτίμηση της σημαντικότητας των δεδομένων.

Η απόδοση της εκπαίδευσης σε μια εποχή εκπαίδευσης μπορεί να εκφραστεί μαθηματικά μέσω της μετρικής της ελάττωσης της συνολικής απώλειας, ως εξής:

$$\Delta L[k] = L(\boldsymbol{w}[k-1]) - L(\boldsymbol{w}[k]) \tag{3.10}$$

Σύμφωνα με το [6], η ελάττωση της συνολικής απώλειας και η νόρμα της παραγώγου συνδέονται με την εξής σχέση:

$$\Delta L[k] = \gamma \|\boldsymbol{g}[k]\|_2^2, \tag{3.11}$$

όπου γ είναι μια παράμετρος που εξαρτάται από την αρχιτεκτονική του βαθέος νευρωνικού δικτύου. Έπειτα, συνδυάζοντας τις εξισώσεις 3.11 και 3.9 μπορούμε να συνάγουμε ότι το τετράγωνο της προσέγγισης της παραγώγου μπορεί να εκτιμήσει την ελάττωση της συνολικής απώλειας, και άρα την απόδοση της εκπαίδευσης για τον τρέχοντα γύρο επικοινωνίας. Συνε-

πώς, έχουμε τη δυνατότητα να ορίσουμε την σημαντικότητα των δεδομένων με τον τρόπο που ακολουθεί.

Ορισμός 3.1. *Η σημαντικότητα του δείγματος $(\boldsymbol{x}_i, y_i)$ που βρίσκεται στη συσκευή n είναι ίσος με το τετράγωνο της εκτιμώμενης νόρμας της παραγώγου:*

$$\sigma_{n,i} = \rho\gamma \left\| \frac{\partial \ell(\Psi(\boldsymbol{x}_i, \boldsymbol{w}), y_i)}{\partial \boldsymbol{x}_i^L} \right\|_2^2 \tag{3.12}$$

Η σημαντικότητα ενός δείγματος δείχνει την ελάττωση του σφάλματος που μπορεί να αποφέρει στην εκπαίδευση του καθολικού μοντέλου. Έτσι, θα ορίσουμε ως ακολούθως την συνάρτηση της ελάττωσης απώλειας της συσκευής n:

$$g_n(B_n) = \sum_{i=1}^{B_n} \sigma_{n,i}, \quad B_n \leq D_n, \tag{3.13}$$

η οποία αντιπροσωπεύει την συνεισφορά της συσκευής n στην ελάττωση της συνολικής απώλειας, όταν σε αυτήν επιλέγονται τα n σημαντικότερα δεδομένα. Θα θεωρήσουμε μια διάταξη στα δεδομένα με σειρά σημαντικότητας, δηλαδή θα ισχύει $\sigma_{n,1} \geq \sigma_{n,2} \geq \cdots \geq \sigma_{n,D_n}$.

Στην παρούσα εργασία η διεργασία μηχανικής μάθησης με την οποία θα ασχοληθούμε είναι η ταξινόμηση εικόνων σε μία από πολλές κλάσεις. Γι'αυτό, θα χρησιμοποιήσουμε για συνάρτηση απώλειας την ευρέως χρησιμοποιούμενη Categorical Cross-Entropy Loss. Το λήμμα που ακολουθεί είναι ιδιαίτερα χρήσιμο για αυτό το σενάριο:

Λήμμα 3.1. *Όταν χρησιμοποιείται η συνάρτηση απώλειας Categorical Cross-Entropy Loss μαζί με συνάρτηση ενεργοποίησης Softmax στην έξοδο, η εκτιμώμενη νόρμα της παραγώγου είναι ίση με την νόρμα της διαφοράς της πρόβλεψης του μοντέλου $\tilde{y}_i$ από την πραγματική τιμή y_i. Δηλαδή:*

$$\left\| \frac{\partial \ell(\tilde{y}_i, y_i)}{\partial \boldsymbol{x}_i^L} \right\|_2 = \|\tilde{y}_i - y_i\|_2$$

Απόδειξη. Με χρήση του κανόνα της αλυσίδας, η παράγωγος της απώλειας ως προς την είσοδο στην ενεργοποίηση του επιπέδου εξόδου x_i^L είναι

$$\frac{\partial \ell(\tilde{y}_i, y_i)}{\partial \boldsymbol{x}_i^L} = \frac{\partial \ell(\tilde{y}_i, y_i)}{\partial \tilde{y}_i} \frac{\partial \tilde{y}_i}{\partial \boldsymbol{x}_i^L} = \underbrace{\frac{\partial L_{CE}(\tilde{y}_i, y_i)}{\partial \tilde{y}_i}}_{1\times C} \underbrace{\frac{\partial S(\boldsymbol{x}_i^L)}{\partial \boldsymbol{x}_i^L}}_{C\times C}$$

όπου

- Η L_{CE} είναι η συνάρτηση απώλειας Categorical Cross-Entropy Loss, που υπολογίζεται ως η εντροπία πάνω στην κατανομή πιθανότητας της ταμπέλας και της πρόβλεψης ως $L_{CE} = \sum_c -p_c \cdot \log(\tilde{p_c})$. Θεωρούμε εν προκειμένω ότι οι ταμπέλες είναι διανύσματα πιθανοτήτων $y = [p_1 \dots p_C]$ (που όταν τα δεδομένα ανήκουν σε μόνο μία κλάση θα έχουν παντού μηδενικές τιμές εκτός από μία θέση που θα έχουν την τιμή 1) και αντίστοιχα οι προβλέψεις $\tilde{y} = [\tilde{p_1} \dots \tilde{p_C}]$. Τότε, η παράγωγος της απώλειας ως προς τις προβλέψεις $\tilde{y}$ είναι ένα $1 \times C$ διάνυσμα (όπου C ο αριθμός των κλάσεων):

$$\frac{\partial L_{CE}(\tilde{y}_i, y_i)}{\partial \tilde{y}_i} = [-\frac{p_1}{\tilde{p_1}} \cdots - \frac{p_C}{\tilde{p_C}}]$$

- Η $S(\overrightarrow{\cdot})$ εκφράζει την συνάρτηση Softmax. Επειδή η Softmax πρόκειται για μια $\mathbb{R}^n \to \mathbb{R}^n$ συνάρτηση, η παράγωγός της εκφράζεται με την μορφή ενός $n \times n$ Ιακωβιανού πίνακα ως εξής [2, 21]:

$$\nabla S(z) = \begin{pmatrix} \frac{\partial S(z)}{\partial z_1} \\ \frac{\partial S(z)}{\partial z_2} \\ \vdots \\ \frac{\partial S(z)}{\partial z_n} \end{pmatrix} = \begin{pmatrix} S(z_i)(1 - S(z_i)) & \text{αν } i = j \\ S(z_i)(-S(z_j)) & \text{αν } i \neq j \end{pmatrix} = \boldsymbol{S}'_{ij}(z)$$

  Στην δική μας περίπτωση λοιπόν, η παράγωγος είναι

$$\frac{\partial S(\boldsymbol{x}_i^L)}{\partial \boldsymbol{x}_i^L} = \begin{pmatrix} \tilde{p_i}(1 - \tilde{p_i}) & \text{αν } i = j \\ \tilde{p_i}(-\tilde{p_j}) & \text{αν } i \neq j \end{pmatrix},$$

  που έχει τη μορφή ενός $C \times C$ Ιακωβιανού πίνακα. Ο πολλαπλασιασμός του διανύσματος με αυτόν τον πίνακα στον κανόνα της αλυσίδας λοιπόν παράγει την παράγωγο που μας ενδιαφέρει σε μορφή ενός $1 \times C$ διανύσματος όπου το $i-$οστό στοιχείο θα είναι ίσο με

$$-p_i \cdot (1 - \tilde{p_i}) + \sum_{j \neq i} p_j \tilde{p_i} = -p_i \cdot (1 - \tilde{p_i}) + \tilde{p_i} \cdot (1 - p_i) = p_i - \tilde{p_i}$$

Επομένως, για το διάνυσμα που αποτελεί την παράγωγο θα ισχύει

$$\frac{\partial \ell(\tilde{y}_i, y_i)}{\partial \boldsymbol{x}_i^L} = y_i - \tilde{y}_i$$

και η εκτιμώμενη νόρμα της παραγώγου θα αποτελεί την Ευκλείδια απόσταση της πρόβλεψης από την ταμπέλα:

$$\left\| \frac{\partial \ell(\tilde{y}_i, y_i)}{\partial \boldsymbol{x}_i^L} \right\|_2 = \|y_i - \tilde{y}_i\|_2$$

■

Το γεγονός ότι η B_n αποτελεί διακριτή μεταβλητή θα καθιστούσε την ανάλυση της συνάρτησης δύσκολη. Ωστόσο, επειδή τυπικά τα σύνολα δεδομένων $\mathcal{D}_n$ έχουν μεγάλο μέγεθος, μπορούμε να χαλαρώσουμε την μεταβλητή B_n και να την θεωρήσουμε συνεχή. Αντίστοιχα, η συνάρτηση $g_n(B_n)$ μπορεί να αναλυθεί σαν μία τμηματικά γραμμική συνάρτηση χρησιμοποιώντας τις τιμές της στα σημεία των ακεραίων B_n, όπως φαίνεται στην γραφική παράσταση του σχήματος 3.2. Δηλαδή, η g_n μπορεί να οριστεί ως εξής:

$$g_n(B_n) = \sigma_{n,\lceil B_n \rceil} \cdot (B_n - \lfloor B_n \rfloor) + \sum_{i=1}^{B_n} \sigma_{n,i}, \;\; 0 \leq B_n \leq D_n, \tag{3.14}$$

όπου οι συναρτήσεις $\lceil\cdot\rceil$ και $\lfloor\cdot\rfloor$ αποτελούν τη στρογγυλοποίηση προς τα πάνω και προς τα κάτω αντίστοιχα.

Τότε, προκύπτει το ακόλουθο λήμμα:

Λήμμα 3.2. *Η $g_n(B_n)$ με συνεχή μεταβλητή B_n αποτελεί κοίλη συνάρτηση.*

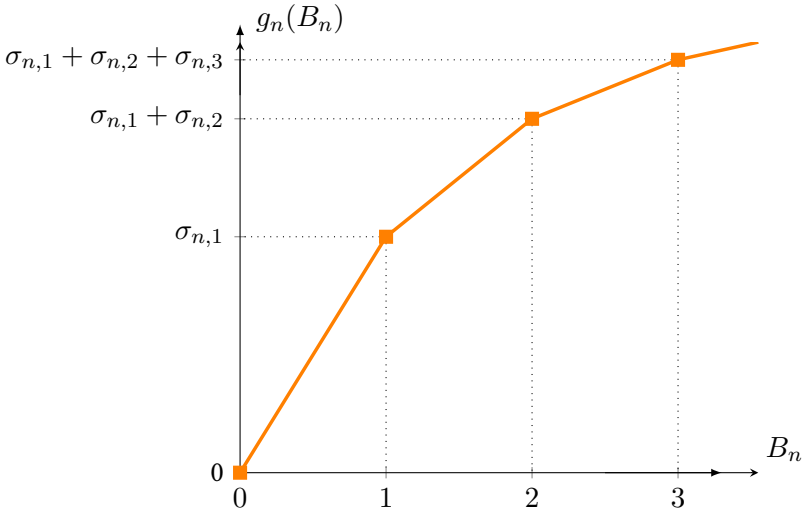
Απόδειξη. Ακολουθώντας το [15], επαναδιατυπώνουμε την συνάρτηση $g_n(B_n)$ ως το ανά σημείο ελάχιστο πολλών γραμμικών συναρτήσεων, ως εξής

$$g_n(B_n) = \min\{h_{n,1}(B_n), h_{n,2}(B_n), \ldots, h_{n,D_n}(B_n)\}, \quad 0 \leq B_n \leq D_n, \tag{3.15}$$

όπου $h_{n,B_m}(B_n) = \sigma_{n,\lceil B_m \rceil} \cdot (B_n - \lfloor B_m \rfloor) + \sum_{i=1}^{B_m} \sigma_{n,i}, \quad 0 \leq B_n \leq D_n$. Σύμφωνα με το [5], το ανά σημείο ελάχιστο $g_n(B_n)$ είναι κοίλη συνάρτηση επειδή οι $h_{n,1}(B_n), h_{n,2}(B_n), \ldots, h_{n,D_n}(B_n)$ είναι όλες γραμμικές. ■

Σημειώνεται ότι η g_n όπως ορίστηκε δεν είναι παραγωγίσιμη σε ορισμένα σημεία, ωστόσο η υποπαράγωγός της (subgradient) μπορεί να εκφραστεί ως

$$\frac{\partial g_n}{\partial B_n} \begin{cases} = \sigma_{n,\lceil B_n \rceil}, & \text{αν } B_n \notin \mathbb{Z} \\ \in [\sigma_{n,B_n+1}, \sigma_{n,B_n}], & \text{αλλιώς} \end{cases} \tag{3.16}$$



Σχήμα 3.2: Ποιοτική γραφική αναπαράσταση της συνάρτησης ελάττωσης της τοπικής απώλειας g_n, με χαλαρωμένη σε συνεχή μεταβλητή B_n και παρεμβολή των τιμών της ως αυτών μιας τμηματικά γραμμικής συνάρτησης.

Η συνάρτηση $g_n(B_n)$, με συνεχή μεταβλητή B_n όπως περιγράφηκε παραπάνω, αποτελεί κοίλη συνάρτηση σύμφωνα με το λήμμα 3.2, γεγονός που θα αποβεί χρήσιμο στην διαδικασία της βελτιστοποίησης. Στηριζόμενοι στα παραπάνω, η συνολική ελάττωση της απώλειας , που θα μας απασχολήσει ως μετρική για την απόδοση της εκπαίδευσης του συστήματος, μπορεί να εκφραστεί ως $\Delta L = \sum_{n \in \mathcal{N}} g_n(B_n)$. Αυτή η μετρική θα ενσωματωθεί στην μοντελοποίηση του συνολικού προβλήματος, αντιπροσωπεύοντας την εκτίμηση του κατά πόσο τα επιλεγμένα δεδομένα με την στρατηγική που περιγράφηκε θα βοηθήσουν στην εκπαίδευση του μοντέλου.

3.3 Υπολογιστικό Μοντέλο

Ακολουθώντας την δουλειά των [36], θα υιοθετήσουμε υπολογιστικό μοντέλο που υποθέτει την ύπαρξη πολλών, πιθανώς ετερογενών επεξεργαστών σε μία συσκευή. Θα υποθέσουμε ότι η κάθε συσκευή είναι εξοπλισμένη με Q_n επεξεργαστές, που καθορίζουν ένα σύνολο $\mathcal{Q}_n = \{1, \dots Q_n\}$. Επειδή η διαχείριση των δεδομένων για την εκπαίδευση αποτελεί διεργασία με αξιόλογο υπολογιστικό φόρτο, θεωρούμε ότι εν γένει ισχύει για το πλήθος επεξεργαστών ανά συσκευή ότι $Q_n \geq 1$.

Ο φόρτος εργασίας σε κάθε επεξεργαστή είναι ανάλογος του συνόλου δεδομένων που καλείται να χειριστεί. Έαν το επιλεγμένο σύνολο δεδομένων μετά την διαδικασία που περιγράφηκε στο υποκεφάλαιο 3.2 είναι το $\mathcal{B}_n$, αυτό κατακερματίζεται ανάμεσα στους επεξεργαστές που αποτελούν το σύνολο $\mathcal{Q}_n$. Έτσι ορίζουμε τα σύνολα $\mathcal{B}_n^q \subseteq \mathcal{B}_n$ που αποτελούν τα σύνολα δεδομένων που έχουν ανατεθεί σε κάθε επεξεργαστή $q \in \mathcal{Q}_n$, με τρόπο τέτοιο ώστε να ισχύει ότι $\bigcup_{q=1}^{Q_n} \mathcal{B}_n^q = \mathcal{B}_n$. Έπεται ότι ο φόρτος εργασίας του επεξεργαστή q της συσκευής n σε FLOPs δίνεται από τον τύπο $W_n^q = B_n^q \cdot N_{FLOPS}$, όπου N_{FLOPS} είναι ο αριθμός των πράξεων κινητής υποδιαστολής που απαιτούνται για τον χειρισμό ενός δείγματος.

Αφού υπολογιστεί η παράγωγος με την οπισθοδιάδοση, ακολουθεί η συνάθροιση των παραγώγων στην πλευρά του διακομιστή, η οποία θεωρείται αμελητέα επειδή στον αλγόριθμο FedAvg αποτελεί έναν απλό υπολογισμό μέσου όρου. Δεδομένου των φόρτων που έχουν να διαχειριστούν οι επεξεργαστές σε μια συσκευή, και των συχνοτήτων λειτουργίας τους, μπορούμε να υπολογίσουμε τον χρόνο που απαιτείται από κάθε επεξεργαστή και συνολικά από την συσκευή για την υπολογιστική διεργασία του γύρου επικοινωνίας της ομόσπονδης μάθησης

$$t_{n,q}^{cmp} = \frac{W_n^q}{f_n^q}, \quad \forall n \in \mathcal{N} \;\; \forall q \in \mathcal{Q}_n \;\; [\text{sec}] \tag{3.17}$$

$$t_n^{cmp} = \max_{q=1}^{Q_n} \left(\frac{W_n^q}{f_n^q} \right), \quad \forall n \in \mathcal{N} \;\; [\text{sec}] \tag{3.18}$$

Θα θεωρήσουμε μία σταθερή ισχύ κατανάλωσης που θα συνεισφέρει στην συνολική ενέργεια κατανάλωσης για τη διάρκεια που ο επεξεργαστής βρίσκεται σε λειτουργία για τη διεργασία υπολογισμού

$$E_{const}^{cmp} = P_{const} \cdot t_{n,q}^{cmp} = P_{const} \frac{W_n^q}{f_n^q} \tag{3.19}$$

Από την δουλειά των [24, 26, 36], μπορούμε να μοντελοποιήσουμε την ενεργειακή κατανάλωση του κάθε επεξεργαστή για τον χειρισμό των δεδομένων συναρτήσει της συχνότητας λειτουργίας του και του φόρτου εργασίας (συνόλου δεδομένων) που αναλαμβάνει - Η ισχύς κατανάλωσης δίνεται από τον τύπο $P_n^q = C_n^q (f_n^q)^3$. Στην ανάλυση υπεισέρχεται η σταθερά C_n^q [Watt(GFLOPs/s)$^{-3}$], που ουσιαστικά εκφράζει την ενεργειακή αποδοτικότητα του επεξεργαστή q της συσκευής n, και ορίζεται ως ο ρυθμός αύξησης της ισχύος ως προς τον κύβο της ταχύτητας του επεξεργαστή σε FLOPs ανά κύκλο ($C_n^q = \Psi_n^q / (\theta_n^q)^3$, όπου το θ_n^q υποδηλώνει τον αριθμό των FLOPs ανά κύκλο στον επεξεργαστή, ενώ το Ψ_n^q αποτελεί μέγεθος εξαρτώμενο από την αρχιτεκτονική του επεξεργαστή που έχει μονάδα μέτρησης τα Watt/(κύκλοι/s)^3). Για τους

σκοπούς μας θα θεωρήσουμε ότι η σταθερά C_n^q εξαρτάται αποκλειστικά από την αρχιτεκτονική του επεξεργαστή. Έτσι, δεδομένου του χρόνου υπολογισμού ενός επεξεργαστή W_n^q/f_n^q και της ισχύος κατανάλωσής του μπορούμε να εκφράσουμε την συνολική ενεργειακή κατανάλωση της συσκευής ως

$$\begin{aligned} E_n^{cmp} &= \sum_{q=1}^{Q_n} \left(C_n^q W_n^q (f_n^q)^2 + P_{const} \frac{W_n^q}{f_n^q} \right) \\ &= \sum_{q=1}^{Q_n} \left(C_n^q B_n^q \cdot N_{FLOPS} (f_n^q)^2 + P_{const} \frac{B_n^q \cdot N_{FLOPS}}{f_n^q} \right), \quad \forall n \in \mathcal{N} \text{ [Joule].} \end{aligned} \tag{3.20}$$

Αυτή θα αποτελέσει βασική μετρική που θα προσπαθήσουμε να ελαττώσουμε κατά την διαδικασία βελτιστοποίησης, εντός των ορίων που καθορίζει και ο χρόνος υπολογισμού.

3.4 Τηλεπικοινωνιακό Μοντέλο

Θεωρούμε δύο στάδια μετάδοσης σε κάθε γύρο επικοινωνίας: Την αποστολή των τοπικών παραγώγων από όλες τις συσκευές στον εξυπηρετητή, και την εκπομπή της συναθροισμένης παραγώγου από τον διακομιστή άκρης προς όλες τις συσκευές, της οποίας την ανάλυση θα παραλείψουμε επειδή δεν επηρεάζεται με κάποιο τρόπο από την διαδικασία βελτιστοποίησης των ραδιοπόρων των συσκευών. Για την ασύρματη επικοινωνία των χρηστών με τον εξυπηρετητή στο στάδιο της αποστολής των τοπικών παραγώγων θα χρησιμοποιηθεί η Μη Ορθογωνική Πολλαπλή Πρόσβαση (NOMA) στο πεδίο της ισχύος.

Η μετάδοση των τοπικών παραγώγων γίνεται ταυτόχρονα σε κανάλι μετάδοσης με εύρος ζώνης B Hz. Συμβολίζουμε με G_n το κέρδος του καναλιού ανάμεσα στη συσκευή n και στον διακομιστή. Αυτό υπολογίζεται με τον τύπο $G_n = \rho \cdot d_n^{-\alpha}$, όπου ρ [dB] είναι η απώλεια διαδρομής σε απόσταση αναφοράς το 1m, d_n είναι η Ευκλείδια απόσταση μεταξύ συσκευής και διακομιστή, και α είναι ο εκθέτης απώλειας διαδρομής. Θεωρούμε χωρίς βλάβη της γενικότητας ότι τα κέρδη του καναλιού είναι ταξινομημένα σε αύξουσα σειρά, ήτοι $G_1 \leq \cdots \leq G_n$, και ότι για την αποκωδικοποίηση των σημάτων χρησιμοποιείται η τεχνική SIC (Successive Interference Cancellation), κατά την οπόια η αποκωδικοποίηση ξεκινάει από τον χρήστη με το μεγαλύτερο κέρδος του καναλιού. Τότε, ο ρυθμός μεταφοράς που μπορεί να επιτευχθεί δίνεται από τον τύπο του Shannon ως

$$R_n = B \log_2 \left(1 + \frac{G_n p_n}{\sum_{i=1}^{n-1} G_i p_i + B \cdot N_0} \right) \text{ [bps]}, \tag{3.21}$$

όπου:

- p_n Watt είναι η ισχύς μετάδοσης της συσκευής n κατά τη ζεύξη ανόδου
- N_0 dBm/Hz είναι η φασματική πυκνότητα ισχύος λευκού γκαουσιανού θορύβου (AWGN) μηδενικού μέσου.

Σημειώνεται επίσης, ότι υποθέτουμε στο πλαίσιο της παρούσας εργασίας ότι υπάρχει τέλεια γνώση για την πληροφορία της κατάστασης του καναλιού (CSI) από την πλευρά των συσκευών· μια υπόθεση που συχνά γίνεται όταν εξετάζονται προβλήματα που εστιάζουν στην ανάθεση ραδιοπόρων.

Με βάση τα παραπάνω, ο απαιτούμενος χρόνος επικοινωνίας για την συσκευή n ορίζεται ως

$$t_n^{tx} = \frac{V}{R_n} \text{ [sec]}, \tag{3.22}$$

όπου V bits αποτελεί το μέγεθος της παραγώγου που πρέπει να μεταδοθεί, το οποίο είναι σταθερό και ίδιο για όλες τις συσκευές.

Το αντίστοιχο ενεργειακό κόστος για την ίδια εργασία υπολογίζεται ως

$$E_n^{tx} = t_n^{tx} \cdot p_n = \frac{V \cdot p_n}{R_n} \text{ [Joule]}. \tag{3.23}$$

Τα δύο αυτά μεγέθη θα χρησιμεύσουν στον καθορισμό της αντικειμενικής συνάρτησης και των περιορισμών του προβλήματος βελτιστοποίησης.

3.5 Μοντελοποίηση του Προβλήματος Βελτιστοποίησης

Έχοντας μοντελοποιήσει α) Την στρατηγική επιλογής δεδομένων και τη μετρική της σημαντικότητας, β) Την ενέργεια και τον χρόνο που απαιτείται για τον υπολογιστικό μέρος της διεργασίας της ομόσπονδης μάθησης σε έναν γύρο επικοινωνίας, και γ) Την ενέργεια που απαιτείται για το κομμάτι της μετάδοσης της παραγώγου κατά την ίδια διεργασία, μπορούμε να ορίσουμε την αντικειμενική συνάρτηση και τους περιορισμούς που θα καθορίσουν το πρόβλημα βελτιστοποίησης με το οποίο θα καταπιαστούμε.

Αρχικά, θα ορίσουμε μια μετρική που θα εκφράζει την ενεργειακή αποδοτικότητα της ομόσπονδης μάθησης σε έναν γύρο επικοινωνίας ως το πηλίκο της σημαντικότητας όλων των επιλεγμένων δεδομένων στο σύστημα προς την συνολική ενεργειακή κατανάλωση για τις διεργασίες του υπολογισμού και της μετάδοσης, όπως περιγράφηκαν στις προηγούμενες υποενότητες. Αυτή η μετρική θα αποτελέσει την αντικειμενική συνάρτηση του προβλήματος, που θα επιχειρήσουμε να μεγιστοποιήσουμε.

$$\begin{aligned} F(\boldsymbol{B_n^q}, \boldsymbol{f_n^q}, \boldsymbol{p_n}) &= \frac{\sum_{n=1}^{N} g_n(B_n)}{E^{cmp} + E^{tx}} \\ &= \frac{\sum_{n=1}^{N} g_n(B_n)}{\left(\begin{array}{r} \sum_{n=1}^{N} \sum_{q=1}^{Q_n} \left(C_n^q B_n^q \cdot N_{FLOPS} (f_n^q)^2 + P_{const} \frac{B_n^q \cdot N_{FLOPS}}{f_n^q} \right) \\ + \sum_{n=1}^{N} \frac{V \cdot p_n}{B \log_2 \left(1 + \frac{G_n p_n}{\sum_{i=1}^{n-1} G_i p_i + N_0} \right)} \end{array} \right)} \end{aligned} \tag{3.24}$$

Έχοντας καθορίσει την αντικειμενική συνάρτηση, και με τις σχέσεις που αναφέρθηκαν κατά τη μοντελοποίηση του συστήματος, μπορούμε να διατυπώσουμε το πρόβλημα βελτιστοποίησης, όπως φαίνεται παρακάτω.

$$\mathcal{P}: \max_{\{B_n, B_n^q, f_n^q, p_n\}} \frac{\sum_{n=1}^{N} g_n(B_n)}{E^{cmp} + E^{tx}} \tag{3.25α΄}$$

$$\text{s.t. } E^{cmp} = \sum_{n=1}^{N}\sum_{q=1}^{Q_n}\left(C_n^q W_n^q (f_n^q)^2 + P_{const}\cdot\frac{B_n^q \cdot N_{FLOPS}}{f_n^q}\right) \tag{3.25β΄}$$

$$E^{tx} = \sum_{n=1}^{N}\frac{V\cdot p_n}{R_n} \tag{3.25γ΄}$$

$$R_n = Blog_2\left(1+\frac{G_n p_n}{\sum_{i=1}^{n-1} G_i p_i + N_0}\right)\ \forall n\in\mathcal{N} \tag{3.25δ΄}$$

$$W_n^q = B_n^q \cdot N_{FLOP},\ \ \forall n\in\mathcal{N},\ \ \forall q\in\mathcal{Q}_n \tag{3.25ε΄}$$

$$f_n^{q,min} \le f_n^q \le f_n^{q,max}\ \ \forall n\in\mathcal{N},\ \ \forall q\in\mathcal{Q}_n \tag{3.25ς΄}$$

$$0\le p_n\le p_n^{max}\ \ \forall n\in\mathcal{N}, \tag{3.25ζ΄}$$

$$0\le B_n\le D_n\ \ \forall n\in\mathcal{N}, \tag{3.25η΄}$$

$$0\le B_n^q\le B_n\ \ \forall n\in\mathcal{N},\ \ \forall q\in\mathcal{Q}_n \tag{3.25θ΄}$$

$$\sum_{q=1}^{Q_n} B_n^q = B_n\ \ \forall n\in\mathcal{N}, \tag{3.25ι΄}$$

$$\frac{V}{R_n}+\frac{B_n^q\cdot N_{FLOP}}{f_n^q}\le T_{max},\ \ \forall n\in\mathcal{N},\forall q\in\mathcal{Q}_n, \tag{3.25ια΄}$$

$$\frac{V}{R_n}\le T_{max}^{tx},\ \ \forall n\in\mathcal{N} \tag{3.25ιβ΄}$$

$$\frac{B_n^q\cdot N_{FLOP}}{f_n^q}\le T_{max}^{cmp},\ \ \forall n\in\mathcal{N},\forall q\in\mathcal{Q}_n \tag{3.25ιγ΄}$$

Χρησιμοποιώντας μόνο τις μεταβλητές βελτιστοποίησης και τις σταθερές του συστήματος, το πρόβλημα διατυπώνεται ισοδύναμα ως ακολούθως:

$$\mathcal{P}1: \max_{\{B_n, B_n^q, f_n^q, p_n\}} \frac{\sum_{n=1}^{N} g_n(B_n)}{\left(\begin{array}{r} \sum_{n=1}^{N} \sum_{q=1}^{Q_n} \left(C_n^q B_n^q \cdot N_{FLOPS} (f_n^q)^2 + P_{const} \frac{B_n^q \cdot N_{FLOPS}}{f_n^q} \right) \\ + \sum_{n=1}^{N} \frac{V \cdot p_n}{B \log_2 \left(1 + \frac{G_n p_n}{\sum_{i=1}^{n-1} G_i p_i + N_0} \right)} \end{array} \right)} \tag{3.26α′}$$

$$\text{s.t. } f_n^{q,min} \leq f_n^q \leq f_n^{q,max} \ \ \forall n \in \mathcal{N}, \ \ \forall q \in \mathcal{Q}_n \tag{3.26β′}$$

$$0 \leq p_n \leq p_n^{max} \ \ \forall n \in \mathcal{N}, \tag{3.26γ′}$$

$$0 \leq B_n \leq D_n \ \ \forall n \in \mathcal{N}, \tag{3.26δ′}$$

$$0 \leq B_n^q \leq B_n \ \ \forall n \in \mathcal{N}, \ \ \forall q \in \mathcal{Q}_n \tag{3.26ε′}$$

$$\sum_{q=1}^{Q_n} B_n^q = B_n \ \ \forall n \in \mathcal{N}, \tag{3.26ς′}$$

$$\frac{V}{B \log_2 \left(1 + \frac{G_n p_n}{\sum_{i=1}^{n-1} G_i p_i + N_0} \right)} + \frac{B_n^q \cdot N_{FLOP}}{f_n^q} \leq T_{max}, \ \ \forall n \in \mathcal{N}, \forall q \in \mathcal{Q}_n \tag{3.26ζ′}$$

$$\frac{V}{B \log_2 \left(1 + \frac{G_n p_n}{\sum_{i=1}^{n-1} G_i p_i + N_0} \right)} \leq T_{max}^{tx}, \ \ \forall n \in \mathcal{N} \tag{3.26η′}$$

$$\frac{B_n^q \cdot N_{FLOP}}{f_n^q} \leq T_{max}^{cmp}, \ \ \forall n \in \mathcal{N}, \forall q \in \mathcal{Q}_n \tag{3.26θ′}$$

4

Επίλυση του Προβλήματος Βελτιστοποίησης

Το πρόβλημα βελτιστοποίησης 3.26 είναι ιδιαιτέρως δύσκολο να επιλυθεί ευθέως, εφόσον τόσο η αντικειμενική συνάρτηση όσο και ο χρονικός περιορισμός δεν είναι κυρτές συναρτήσεις ως προς τις μεταβλητές βελτιστοποίησης. Σε αυτό το κεφάλαιο θα χρησιμοποιήσουμε συνδυαστικά τις τεχνικές κλασικής βελτιστοποίησης που αναφέρθηκαν στο κεφάλαιο 2 για να μετατρέψουμε το πρόβλημα σε μια πιο διαχειρίσιμη μορφή και θα σκιαγραφήσουμε έναν αλγόριθμο που θα οδηγήσει σε μια υποβέλτιστη λύση για το πρόβλημα ταυτόχρονης ανάθεσης πόρων και επιλογής δεδομένων.

4.1 Διάσπαση του Προβλήματος σε Κυρτά Υποπροβλήματα

4.1.1 Μετατροπή του Κλασματικού Προβλήματος Βελτιστοποίησης με την τεχνική του Dinkelbach

Εφόσον η αντικειμενική συνάρτηση αποτελεί πηλίκο δύο θετικών συναρτήσεων, μπορούμε να επαναδιατυπώσουμε το πρόβλημα $\mathcal{P}1$ ως εξής:

$$\max_{\{B_n, B_n^q, f_n^q, p_n\}} \sum_{n=1}^{N} g_n(B_n) - y \cdot \left(\sum_{n=1}^{N} \sum_{q=1}^{Q_n} \left(C_n^q B_n^q \cdot N_{FLOP} \cdot (f_n^q)^2 + P_{const} \frac{B_n^q \cdot N_{FLOPS}}{f_n^q} \right) + \sum_{n=1}^{N} \frac{V \cdot p_n}{B \log_2 \left(1 + \frac{G_n p_n}{\sum_{i=1}^{n-1} G_i p_i + N_0} \right)} \right)$$

$$\text{s.t. } (3.26\alpha') - (3.26\theta') \tag{4.1}$$

Για αυτόν τον σκοπό εισάγουμε την βοηθητική μεταβλητή y, η οποία θα ανανεώνεται επαναληπτικά σύμφωνα με τον τύπο

$$y[t+1] = F(\boldsymbol{B_n^q}[t], \boldsymbol{f_n^q}[t], \boldsymbol{p_n}[t]), \tag{4.2}$$

όπου $F(\boldsymbol{B_n^q}[t], \boldsymbol{f_n^q}[t], \boldsymbol{p_n}[t])$ η τιμή της αντικειμενικής συνάρτησης 3.24 που έχει προκύψει μετά την βελτιστοποίηση των μεταβλητών απόφασης χρησιμοποιώντας ως παράμετρο Dinkelbach

την τιμή $y[t]$. Αυτή η διαδικασία επαναλαμβάνεται μέχρι τη σύγκλιση της τιμής y, οπότε προκύπτει και μια λύση στο πρόβλημά μας.

4.1.2 Εναλλασσόμενη Βελτιστοποίηση

Αξιοποιώντας τον αλγόριθμο του Dinkelbach η αντικειμενική συνάρτηση βρίσκεται σε μια πιο διαχειρίσιμη μορφή, όμως το πρόβλημα εξακολουθεί να είναι δύσκολο να επιλυθεί ευθέως. Οι μεταβλητές βελτιστοποίησης εξακολουθούν να είναι μεταξύ τους συνδεδεμένες στην αντικειμενική συνάρτηση και στους περιορισμούς, με αποτέλεσμα το πρόβλημα να παραμένει μη κυρτό. Για αυτόν τον σκοπό, θα διασπάσουμε το πρόβλημα βελτιστοποίησης σε υποπροβλήματα, με σκοπό αυτό να οδηγήσει μαζί με την μετατροπή Dinkelbach στο να προκύψουν κυρτά υποπροβλήματα, τα οποία θα μπορούν να επιλυθούν ευθέως.

Χρησιμοποιώντας την εναλλασσόμενη βελτιστοποίηση, θα κατατμήσουμε το σύνολο των μεταβλητών απόφασης σε έναν αριθμό υποσυνόλων, και σε κάθε επανάληψη θα ορίζεται το σχετικό υποπρόβλημα ως το αρχικό, κρατώντας όμως όλες τις μεταβλητές εκτός από αυτές ενός συνόλου σταθερές. Τα υποπροβλήματα δηλαδή θα επιλύονται διαδοχικά ως προς ένα μόνο υποσύνολο των μεταβλητών διαδοχικά μέχρι να ικανοποιηθεί κάποιο κριτήριο σύγκλισης (π.χ. σύγκλιση στην τιμή της αντικειμενικής συνάρτησης), οπότε θα θεωρηθεί ότι το συνολικό πρόβλημα έχει καταλήξει σε μία υποβέλτιστη λύση.

Τα υποσύνολα των μεταβλητών, και άρα τα υποπροβλήματα που θα ορίσουμε θα είναι τα εξής:

- Το σύνολο των $\{B_n^q\} \cup \{B_n\}$ για το υποπρόβλημα της ανάθεσης δεδομένων
- Το σύνολο των $\{f_n^q\}$ για το υποπρόβλημα της ανάθεσης υπολογιστικών πόρων
- Το σύνολο των $\{p_n\}$ για το υποπρόβλημα της ανάθεσης ραδιοπόρων

4.2 Επίλυση του Προβλήματος Ανάθεσης Δεδομένων

Αν κρατήσουμε σταθερές όλες τις μεταβλητές εκτός από τις $\{B_n^q\}$ και $\{B_n\}$, που αφορούν το πρόβλημα τις ανάθεσης δεδομένων στις συσκευές και στους επεξεργαστές, τότε ερχόμαστε αντιμέτωποι με το παρακάτω πρόβλημα:

$$\mathcal{P}2: \max_{\{B_n, B_n^q\}} \sum_{n=1}^{n} g(B_n) - y\left(\sum_{n=1}^{N}\sum_{q=1}^{Q_n}\left(C_n^q B_n^q \cdot N_{FLOP} \cdot (f_n^q)^2 + P_{const}\frac{B_n^q \cdot N_{FLOPS}}{f_n^q}\right)\right) \tag{4.3α΄}$$

$$\text{s.t. } B_n^q \frac{N_{FLOP}}{f_n^q} \leq T_{nq}^{cmp}, \ \ \forall n \in \mathcal{N}, \forall q \in \mathcal{Q}_n \tag{4.3β΄}$$

$$0 \leq B_n^q \leq B_n, \ \ \forall n \in \mathcal{N}, \forall q \in \mathcal{Q}_n \tag{4.3γ΄}$$

$$\sum_{q=1}^{Q_n} B_n^q = B_n \ \ \forall n \in \mathcal{N}, \tag{4.3δ΄}$$

όπου $T_{nq}^{cmp} = \min(T_{max} - V/R_n,\ \ T_{max}^{cmp})\ \forall n \in \mathcal{N}, \forall q \in \mathcal{Q}_n$

Αυτό το πρόβλημα περιέχει μία αντικειμενική συνάρτηση που συνιστά ένα άθροισμα κοίλων συναρτήσεων $g_n(B_n)$ (βλ. λήμμα 3.2) από την οποία αφαιρούνται όροι οι οποίοι είναι γραμμικοί ως προς B_n^q. Επίσης, οι περιορισμοί είναι όλοι γραμμικοί ως προς B_n και B_n^q. Το πρόβλημα $\mathcal{P}2$ είναι λοιπόν κυρτό, και μπορεί να επιλυθεί με την μέθοδο των πολλαπλασιαστών Lagrange.

Εισάγοντας τους πολλαπλασιαστές Lagrange λ_n για τις ισότητες 4.3ε' και τους $\mu_{n,q}$ για τις ανισότητες 4.3β', ορίζουμε την μερική συνάρτηση Lagrange:

$$\begin{aligned}\mathcal{L} = -\sum_{n=1}^{n} g(B_n) + y\left(\sum_{n=1}^{N}\sum_{q=1}^{Q_n}\left(C_n^q B_n^q \cdot N_{FLOP}\cdot (f_n^q)^2 + P_{const}\frac{B_n^q \cdot N_{FLOPS}}{f_n^q}\right)\right) \\ + \sum_{n=1}^{N}\lambda_n((\sum_{q=1}^{Q_n} B_n^q) - B_n) \\ + \sum_{n=1}^{N}\sum_{q=1}^{Q_n}\mu_{n,q}\left(B_n^q\frac{N_{FLOP}}{f_n^q} - T_{nq}^{cmp}\right),\end{aligned} \tag{4.4}$$

και με βάση αυτήν η βέλτιστη λύση ως προς B_n, B_n^q μπορεί να προκύψει μέσω των συνθηκών Karush-Kuhn-Tucker, ως εξής:

$$\frac{\partial\mathcal{L}}{\partial B_n^{q*}} = y\left(C_n^q N_{FLOP}(f_n^q)^2 + P_{const}\frac{N_{FLOPS}}{f_n^q}\right) - \lambda_n^* + \mu_{n,q}^*\frac{N_{FLOP}}{f_n^q} = 0,\ \ 0 \le B_n^{q*} \le D_n \tag{4.5}$$

$$\frac{\partial\mathcal{L}}{\partial B_n^*} = -\frac{\partial g_n(B_n^*)}{\partial B_n^*} + \lambda_n^* = 0 \to \lambda_n^* = \frac{\partial g_n(B_n^*)}{\partial B_n^*},\ \ 0 \le B_n^* \le D_n \tag{4.6}$$

$$\mu_{nq}^*\cdot\left(\frac{B_n^{q*}\cdot N_{FLOP}}{f_n^q} - T_{nq}^{cmp}\right) = 0 \to \mu_{n,q}^* = 0\ \vee B_n^{q*} = \frac{f_n^q}{N_{FLOP}}T_{nq}^{cmp} \tag{4.7}$$

$$\frac{\partial\mathcal{L}}{\partial\lambda_n^*} = 0 \to B_n^* = \sum_{q=1}^{Q_n} B_n^{q*} \tag{4.8}$$

Θα εστιάσουμε αρχικά στην σχέση 4.7, από την οποία διακρίνουμε δύο περιπτώσεις. Στην πρώτη περίπτωση, ο πολλαπλασιαστής $\mu_{n,q}$ είναι θετικός, οπότε ο χρονικός περιορισμός είναι ενεργός, ήτοι η σχετική ανισότητα ικανοποιείται με ισότητα. Ουσιαστικά, δεν περισσεύει χρόνος για την υπολογιστική διεργασία. Σε αυτήν την περίπτωση μπορούμε να βρούμε αμέσως την βέλτιστη τιμή B_n^{q*} για την ανάθεση δεδομένων στον επεξεργαστή (n, q), όπως φαίνεται από την ίδια σχέση.

Στην άλλη περίπτωση, ο χρονικός περιορισμός για τον επεξεργαστή (n, q) δεν είναι ενεργός, δηλαδή η ανίσωση ικανοποιείται χωρίς να ισχύει η ισότητα, και ο πολλαπλασιαστής $\mu_{n,q}^*$ είναι μηδέν. Τότε δεν προκύπτει ευθέως η λύση για την ανάθεση δεδομένων στον συγκεκριμένο επεξεργαστή. Ωστόσο, συνδυάζοντας το γεγονός ότι $\mu_{n,q}^* = 0$ με τις σχέσεις 4.5 και 4.6, προκύπτει η παρακάτω σχέση:

$$\frac{\partial g_n(B_n^*)}{\partial B_n^*} = y\left(C_n^q N_{FLOP}(f_n^q)^2 + P_{const}\frac{N_{FLOPS}}{f_n^q}\right) \tag{4.9}$$

και επιλύοντας αυτήν την εξίσωση ως προς B_n^* προκύπτει η βέλτιστη λύση για την ανάθεση δεδομένων στην συσκευή n. Αυτό στην ουσία σημαίνει ότι στην συσκευή n υπάρχει διαθέσιμος χρόνος για να ανατεθούν περισσότερα από B_n^* δεδομένα, όμως αυτό δεν συμφέρει να γίνει από την άποψη ότι δεν οδηγεί στην μεγιστοποίηση της αντικειμενικής συνάρτησης που συνυπολογίζει την σημαντικότητα των επιλεγμένων δεδομένων και την ενέργεια που καταναλώνεται για την εκπαίδευση σε αυτά.

Για να καταλήξουμε στην εύρεση της λύσης για όλες τις συσκευές και τους επεξεργαστές αρκεί να παρατηρήσουμε επιπλέον ότι αν θεωρήσουμε το πρόβλημα κατά το οποίο θέλουμε να αναθέσουμε ένα δοσμένο πλήθος δεδομένων σε μια συσκευή ύπο τον χρονικό περιορισμό του υπολογισμού, με σκοπό να ελαχιστοποιήσουμε την ενέργεια υπολογισμού, τότε παρατηρούμε ότι θεωρώντας τις B_n, B_n^q συνεχείς μεταβλητές όπως αναφέρθηκε στο προηγούμενο κεφάλαιο, το πρόβλημα μπορεί να αναχθεί στο γνωστό πρόβλημα του κλασματικού σακιδίου (Fractional Knapsack) [12]. Η "αξία" των "αντικειμένων"-επεξεργαστών αντιπροσωπεύεται από τους όρους $C_n^q \cdot (f_n^q)^2$ που εκφράζουν την συγκριτική ενεργειακή κατανάλωση υπό την συχνότητα λειτουργίας τους, ενώ το κόστος τους από τον χρόνο επεξεργασίας ανάλογα με το πλήθος των δεδομένων.

Έχοντας αυτό κατά νου συμπεραίνουμε ότι το πρόβλημα μπορεί να επιλυθεί με άπληστο τρόπο, αν σε κάθε συσκευή κατατάξουμε τους επεξεργαστές σύμφωνα με τους όρους $C_n^q \cdot (f_n^q)^2$, και έπειτα αναθέσουμε όσα περισσότερα δεδομένα μπορούμε σε κάθε επεξεργαστή χωρίς να ξεπεράσουμε το χρονικό όριο.

Τελικά, για να βρούμε τη λύση στο συνολικό πρόβλημα της ανάθεσης δεδομένων πρέπει να συνδυάσουμε τη δομή του Fractional Knapsack που παρατηρείται σε αυτό με την λύση της εξίσωσης 4.9 από την οποία προκύπτει το πλήθος των δεδομένων που πρέπει να αναθέσουμε στη συσκευή για να μεγιστοποιηθεί η τιμή της αντικειμενικής συνάρτησης. Προτείνεται λοιπόν η εξής διαδικασία: Για κάθε επεξεργαστή, ξεκινώντας από τον πιο ενεργειακά αποδοτικό, συγκρίνουμε τις λύσεις B_n^q για τις περιπτώσεις όπου α) Ο χρονικός περιορισμός είναι ενεργός, δηλ. ισχύει η σχέση για το B_n^q στην 4.5, και β) Ο χρονικός περιορισμός δεν είναι ενεργός, οπότε ισχύει η λύση από την σχέση 4.9. Η λύση που θα κρατήσουμε θα είναι η μικρότερη από τις δύο. Αν αυτή είναι η α), τότε θα προχωρήσουμε στον επόμενο επεξεργαστή και θα συνεχίσουμε παρομοίως. Αν ωστόσο η λύση β) είναι μικρότερη, τότε συμπεραίνουμε ότι για τον παρών επεξεργαστή συμφέρει, όσον αφορά την σημαντικότητα/ενέργεια, να αναθέσουμε λιγότερα δεδομένα από τα μέγιστα που μπορούμε λόγω του χρονικού ορίου. Επομένως, στους υπόλοιπους επεξεργαστές, που είναι λιγότερο αποδοτικοί, συμφέρει να μην αναθέσουμε κανένα δεδομένο, δηλαδή να τους κρατήσουμε κλειστούς για αυτόν τον γύρο επικοινωνίας.

Αυτή η διαδικασία σκιαγραφείται στον αλγόριθμο 1, που περιγράφει την στρατηγική επιλογής δεδομένων που προτείνεται στα πλαίσια αυτής της εργασίας.

Αλγόριθμος 1 Βελτιστοποίηση Ανάθεσης Δεδομένων στους Επεξεργαστές μίας Συσκευής

$B_n \leftarrow 0$ {Πλήθος δειγμάτων που έχουν ανατεθεί ως τώρα στη συσκευή}

Για κάθε επεξεργαστή $q \in \{1, \ldots, Q_n\}$

$B^q_{n,Time} \leftarrow \frac{f^q_n}{N_{FLOP}}(T_{max} - \frac{V}{R_n})$

Υπολόγισε το πλήθος $B^q_{n,Energy}$ σύμφωνα με την λύση της εξίσωσης 4.9 ως προς B^q_n, αντικαθιστώντας στον τύπο $B^*_n = B_n + B^q_n$

Εάν $B^q_{n,Time} < B^q_{n,Energy}$

$B^q_n \leftarrow B^q_{n,Time}$

$B_n \leftarrow B_n + B^q_n$

Αλλιώς

$B^q_n \leftarrow B^q_{n,Energy}$

$B_n \leftarrow B_n + B^q_n$

$B^{q'}_n \leftarrow 0$ για όλους τους εναπομείναντες επεξεργαστές $q' \in \{q+1, \ldots, Q_n\}$

Επίστρεψε $B_n, \{B^q_n\}$

Τέλος Επανάληψης

Επίστρεψε $B_n, \{B^q_n\}$

4.3 Επίλυση του Προβλήματος Ανάθεσης Υπολογιστικών Πόρων

Για να καθορίσουμε τις συχνότητες στις οποίες θα ρυθμίσουμε την λειτουργία των επεξεργαστών, θα θεωρήσουμε το σχετικό υποπρόβλημα ως προς τις μεταβλητές $\{f^q_n\}$ κρατώντας όλες τις υπόλοιπες σταθερές. Τότε από το πρόβλημα $\mathcal{P}1$ (3.26) προκύπτει το παρακάτω υποπρόβλημα:

$$\mathcal{P}3: \min_{\{f^q_n\}} \sum_{n=1}^{N}\sum_{q=1}^{Q_n}\left(C^q_n B^q_n \cdot N_{FLOP} \cdot (f^q_n)^2 + P_{const}\frac{B^q_n \cdot N_{FLOPS}}{f^q_n}\right) \quad (4.10\alpha')$$

$$\text{s.t. } B^q_n \cdot N_{FLOP} \leq f^q_n \cdot T^{cmp}_n, \ \ \forall n \in \mathcal{N}, \ \ \forall q \in \mathcal{Q}_n \quad (4.10\beta')$$

$$f^{q,min}_n \leq f^q_n \leq f^{q,max}_n, \ \ \forall n \in \mathcal{N}, \ \ \forall q \in \mathcal{Q}_n, \quad (4.10\gamma')$$

όπου $T^{cmp}_n = \min\left(T_{max} - \frac{V}{R_n}, \ \ T^{cmp}_{max}\right)$.

Ο χρονικός περιορισμός μέσα από απλούς αλγεβρικούς χειρισμούς έχει μετατραπεί σε γραμμικό ως προς f^q_n, ενώ η αντικειμενική συνάρτηση είναι κυρτή ως προς κάθε f^q_n εφόσον αποτελεί άθροισμα κυρτών συναρτήσεων. Το πρόβλημα $\mathcal{P}3$ μπορεί να επιλυθεί λοιπόν με την μέθοδο των πολλαπλασιαστών Lagrange. Θεωρούμε την μερική Λαγκραντζιανή συνάρτηση με τους πολλαπλασιαστές $\mu_{nq}, \ \ \forall n \in \mathcal{N}, \ \ \forall q \in \mathcal{Q}_n$

$$\mathcal{L} = \sum_{n=1}^{N}\sum_{q=1}^{Q_n}\left(C_n^q B_n^q \cdot N_{FLOP} \cdot (f_n^q)^2 + P_{const}\frac{B_n^q \cdot N_{FLOPS}}{f_n^q}\right) + \sum_{n=1}^{N}\sum_{q=1}^{Q_n}\mu_{nq}\left(B_n^q \cdot (N_{FLOPS}) - f_n^q \cdot T_n^{cmp}\right) \tag{4.11}$$

Για άλλη μια φορά θα χρησιμοποιήσουμε τις συνθήκες ΚΚΤ. Για αρχή μπορούμε να παρατηρήσουμε ότι οι συχνότητες επεξεργασίας δεν συνδέονται κάπως μεταξύ τους, οπότε μπορούμε να εξετάσουμε ξεχωριστά το πρόβλημα για κάθε επεξεργαστή. Μπορούμε πρώτα να θεωρήσουμε την περίπτωση που για έναν επεξεργαστή (n, q) ο χρονικός περιορισμός 4.10β΄ θα είναι ενεργός, δηλαδή ισχύει με ισότητα, και άρα θα ισχύει ότι $\mu_{n,q}^* > 0$ από τη συνθήκη του complementary slackness. Βέβαια λόγω της ισότητας θα προκύπτει αμέσως η λύση για την συχνότητα για να ικανοποιείται ο χρονικός περιορισμός, που θα έχει την τιμή (συνυπολογίζοντας και τα όρια συχνότητας)

$$f_n^{q,Time} = \max\left(\frac{B_n^q \cdot N_{FLOP}}{T_n^{cmp}},\ \ f_n^{q,min}\right) \tag{4.12}$$

Στην περίπτωση που ο περιορισμός δεν είναι ενεργός για τον επεξεργαστή (n, q), τότε από τη συνθήκη complementary slackness θα ισχύει ότι $\mu_{nq} = 0$, και συνδυάζοντας αυτό με την συνθήκη ΚΚΤ της στασιμότητας έχουμε

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial f_n^{q*}} = 0 &\rightarrow 2C_n^q B_n^q N_{FLOPS} f_n^{q*} - P_{const}\frac{B_n^q \cdot N_{FLOPS}}{(f_n^{q*})^2} = 0 \\ &\rightarrow (f_n^{q*})^3 = \frac{P_{const}}{2C_n^q}\end{aligned} \tag{4.13}$$

Από αυτήν την σχέση αποκτούμε την συχνότητα λειτουργίας που οδηγεί στην χαμηλότερη κατανάλωση ενέργειας στην περίπτωση που δεν ισχύει ο χρονικός περιορισμός. Δηλαδή θεωρούμε την "ενεργειακά αποδοτική" συχνότητα λειτουργίας

$$f_n^{q,Energy} = \left[\sqrt[3]{\frac{P_{const}}{2C_n^q}}\right]_{f_n^{q,min}}^{f_n^{q,max}}, \tag{4.14}$$

όπου χρησιμοποιούμε τον συμβολισμό $[x]_a^b = \max(a, \min(x, b))$ για να τοποθετήσουμε την ποσότητα x μέσα στα όρια $[a, b]$.

Τελικά, η βέλτιστη λύση για κάθε συχνότητα λειτουργίας f_n^q προκύπτει επιλέγοντας είτε την λύση $f_n^{q,Energy}$ που εξασφαλίζει την ελάχιστη κατανάλωση ενέργειας, αρκεί να ικανοποιείται ο χρονικός περιορισμός (που συμβαίνει αν $f_n^{q,Energy} > f_n^{q,Time}$), είτε την λύση $f_n^{q,Time}$, στην περίπτωση που η πρώτη δεν ικανοποιεί τον περιορισμό. Η λύση $f_n^{q,Time}$ είναι η χαμηλότερη συχνότητα για την οποία ικανοποιούνται όλοι οι περιορισμοί του προβλήματος, και άρα η κοντινότερη στην $f_n^{q,Energy}$ αν αυτή δεν είναι εφικτή. Δηλαδή, η συχνότητα που τελικά επιλέγουμε είναι η

$$f_n^{q*} = \max(f_n^{q,Time}, f_n^{q,Energy}) \tag{4.15}$$

4.4 Επίλυση του Προβλήματος Ανάθεσης Ραδιοπόρων

Για το τρίτο υποπρόβλημα θα επιχειρήσουμε με παρόμοιο τρόπο με πριν να βελτιστοποιήσουμε τις ισχύς μετάδοσης p_n προκειμένου το στάδιο της αποστολής της παραγώγου από κάθε συσκευή να γίνει εντός των χρονικών ορίων και με τη λιγότερη δυνατή ενέργεια. Κρατώντας πάλι όλες τις μεταβλητές σταθερές εκτός από τις p_n, ερχόμαστε αντιμέτωποι με το κάτωθι πρόβλημα:

$$\mathcal{P}4: \min_{\{p_n\}} \sum_{n=1}^{N} p_n t_n^{tx} \tag{4.16α΄}$$

$$\text{s.t. } 0 \leq p_n \leq p_n^{max} \ \ \forall n \in \mathcal{N}, \tag{4.16β΄}$$

$$t_n^{tx} \leq T_n^{tx}, \ \ \forall n \in \mathcal{N}, \forall q \in \mathcal{Q}_n, \tag{4.16γ΄}$$

$$t_n^{tx} = \frac{V}{B \log_2 \left(1 + \frac{G_n p_n}{\sum_{i=1}^{n-1} G_i p_i + N_0}\right)}, \tag{4.16δ΄}$$

όπου $T_n^{tx} = \min\left(\max_{q \in \mathcal{Q}_n}\left(T_{max} - \frac{B_n^q \cdot N_{FLOP}}{f_n^q}\right), \ T_{max}^{tx}\right)$.

Αυτό το πρόβλημα είναι δύσκολο να βελτιστοποιηθεί ευθέως, καθώς η αντικειμενική συνάρτηση είναι μη κυρτή, ενώ ιδιαίτερη πρόκληση αποτελεί το γεγονός ότι η ισχύς μετάδοσης μιας συσκευής επηρεάζει την ενέργεια και τον χρόνο μετάδοσης όχι μόνο της ίδιας, αλλά και όλων των υπολοίπων συσκευών με μεγαλύτερο κέρδος καναλιού -στις οποίες παρεμβάλλεται- λόγω της τεχνικής SIC. Ωστόσο, μπορούμε να απλοποιήσουμε το πρόβλημα όπως γίνεται και στην δουλειά των [50]: Αξιοποιούμε το γεγονός ότι ο χρόνος μετάδοσης t_n^{tx} είναι φραγμένος από τον T_n^{tx}, οπότε μπορούμε να τον παραλείψουμε από το πρόβλημα βελτιστοποίησης χωρίς να έχουμε κάποια σημαντική απώλεια στην τιμή της αντικειμενικής συνάρτησης κατά την επίλυση. Έτσι, προκύπτει ένα πρόβλημα ελαχιστοποίησης του αθροίσματος των ισχύων, και με απλούς αλγεβρικούς χειρισμούς πάλι μετατρέπουμε τον χρονικό περιορισμό σε γραμμικό ως προς p_n. Πρόκειται λοιπόν για ένα γραμμικό πρόγραμμα, και μπορεί να επιλυθεί πολύ πιο εύκολα από το 4.16, για το οποίο θα έπρεπε να καταφύγουμε σε πολύ πιο υπολογιστικά κοστοβόρες τεχνικές.

$$\mathcal{P}4: \min_{p_n} \sum_{n=1}^{N} p_n \tag{4.17α΄}$$

$$\text{s.t. } 0 \leq p_n \leq p_n^{max} \ \ \forall n \in \mathcal{N}, \tag{4.17β΄}$$

$$G_n p_n \geq \left(2^{V/(B \cdot T_n^{tx})} - 1\right) \cdot \left(N_0 B + \sum_{i=1}^{n-1} G_i p_i\right), \ \ \forall n \in \mathcal{N} \tag{4.17γ΄}$$

Για να επιλύσουμε αυτό το πρόβλημα βελτιστοποίησης αρκεί να θέσουμε ξεχωριστά την κάθε ισχύ στην ελάχιστη δυνατή τιμή της ωστέ να ικανοποιείται ο χρονικός περιορισμός, ξεκινώντας από την συσκευή με το μικρότερο κέρδος καναλιού. Με αυτόν τον τρόπο μάλιστα

ελαχιστοποιούμε την παρεμβολή της συσκευής στα σήματα των υπολοίπων συσκευών με μεγαλύτερο κέρδος καναλιού. Βελτιώνεται έτσι και ο ρυθμός μεταφοράς, μειώνοντας τον απαιτούμενο χρόνο και την συνολική ενέργεια.

Η λύση λοιπόν στο προσεγγιστικό/απλοποιημένο υποπρόβλημα βελτιστοποίησης των ισχύων μετάδοσης δίνεται από τις παρακάτω σχέσεις:

$$
\begin{aligned}
&G_1 p_1 = \left(2^{V/(B \cdot T_1^{tx})} - 1\right) \cdot N_0 B \\
&G_2 p_2 = \left(2^{V/(B \cdot T_2^{tx})} - 1\right) \cdot (N_0 B + G_1 p_1) \\
&\vdots \\
&G_n p_n = \left(2^{V/(B \cdot T_n^{tx})} - 1\right) \cdot \left(N_0 B + \sum_{i=1}^{n-1} G_i p_i\right), \quad \forall n \in \mathcal{N}
\end{aligned} \tag{4.18}
$$

4.5 Επίλυση του Συνολικού Προβλήματος Βελτιστοποίησης

Έχοντας επιλύσει τα τρία υποπροβλήματα της ανάθεσης δεδομένων, υπολογιστικών πόρων, και ραδιοπόρων θα προχωρήσουμε στην διατύπωση της διαδικασίας εύρεσης μιας λύσης για το συνολικό πρόβλημα.

Η βασική ιδέα της εναλλασσόμενης βελτιστοποίησης έγκειται στην επαναληπτική επίλυση κάθε ξεχωριστού υποπροβλήματος μέχρι να υπάρξει σύγκλιση σε μία υποβέλτιστη λύση για το συνολικό πρόβλημα. Ωστόσο δεν πρέπει να ξεχάσουμε ότι για να φτάσουμε στις λύσεις των υποπροβλημάτων (και συγκεκριμένα αυτό της ανάθεσης δεδομένων) χρησιμοποιήσαμε τη μέθοδο του Dinkelbach. Επομένως, η διαδικασία εύρεσης της συνολικής λύσης θα πρέπει να ενσωματωθεί στον αλγόριθμο του Dinkelbach, και να επαναλαμβάνεται με κάθε ανανέωση στην παράμετρο Dinkelbach y. Ουσιαστικά ο αλγόριθμος Dinkelbach σε κάθε επανάληψη θα τροφοδοτεί τον αλγόριθμο εύρεσης συνολικής λύσης με την νέα παράμετρο y και ανάποδα, η παράμετρος y θα υπολογίζεται με βάση τη λύση που δίνει ο αλγόριθμος. Αυτό θα επαναλαμβάνεται μέχρι να υπάρξει σύγκλιση.

Η διαδικασία αυτή βελτιστοποίησης των μεταβλητών του συστήματος ομόσπονδης μάθησης για έναν γύρο επικοινωνίας διατυπώνεται ακολούθως και σε μορφή ψευδοκώδικα, στους αλγορίθμους 2 και 3.

Αλγόριθμος 2 Αλγόριθμος Dinkelbach για τη Συνολική Βελτιστοποίηση του $\mathcal{P}1(3.26)$

$y \leftarrow 0$ {Αρχικοποίηση της παραμέτρου Dinkelbach}

Επανάλαβε

Βρες λύση στο πρόβλημα βελτιστοποίησης 4.1 με την τρέχουσα τιμή για το y, χρησιμοποιώντας τον αλγόριθμο 3

Ανανέωσε την τιμή του y σύμφωνα με την εξίσωση 4.2, χρησιμοποιώντας τις τιμές των B_n^q, f_n^q, p_n που προέκυψαν κατά το προηγούμενο βήμα

Εώς ότου υπάρξει σύγκλιση στο y

Επίστρεψε τις τελικές βελτιστοποιημένες τιμές των B_n, B_n^q, f_n^q, p_n

Αλγόριθμος 3 Βελτιστοποίηση Ανάθεσης Δεδομένων, Ραδιοπόρων, και Υπολογιστικών Πόρων στο Σύστημα Ομόσπονδης Μάθησης (Εναλλασσόμενη Βελτιστοποίηση)

{Η βελτιστοποίηση εκτελείται για δεδομένη τιμή της παραμέτρου Dinkelbach y}

Αρχικοποίησε τις μεταβλητές B_n, B_n^q, f_n^q, p_n ώστε να βρεθεί μία εφικτή λύση στο πρόβλημα βελτιστοποίησης 4.1

Για πλήθος επαναλήψεων $i \in \{1, \ldots, \text{maxIter}\}$ (ή μέχρι να υπάρξει σύγκλιση)

Ανανέωσε τις τιμές των μεταβλητών B_n, B_n^q με βάση τον αλγόριθμο 1, χρησιμοποιώντας τις πιο πρόσφατες τιμές για τις μεταβλητές f_n^q, p_n.

Ανανέωσε τις τιμές των μεταβλητών f_n^q με βάση την εξίσωση 4.15, χρησιμοποιώντας τις πιο πρόσφατες τιμές για τις μεταβλητές B_n, B_n^q, p_n.

Ανανέωσε τις τιμές των μεταβλητών p_n με βάση τις εξισώσεις 4.18, χρησιμοποιώντας τις πιο πρόσφατες τιμές για τις μεταβλητές B_n, B_n^q, f_n^q.

Τέλος Επανάληψης

Επίστρεψε Τις τελικές τιμές των μεταβλητών B_n, B_n^q, f_n^q, p_n

5 Διεξαγωγή Πειραμάτων

Έχοντας διατυπώσει τη προτεινόμενη προσέγγιση στο πρόβλημα της κοινής ανάθεσης δεδομένων, ραδιοπόρων και υπολογιστικών πόρων στο σύστημα της ομόσπονδης μάθησης, θα προχωρήσουμε στην ανάπτυξη ενός συστήματος ομόσπονδης μάθησης για να αξιολογήσουμε μέσω μιας σειράς πειραμάτων την απόδοση τόσο κάθε ξεχωριστού συστατικού της προτεινόμενης λύσης όσο και της συνολικής διαδικασίας επίλυσης. Σε γενικές γραμμές θα επιχειρήσουμε να αποτιμήσουμε

- Κατά πόσο βελτιστοποιείται το πρόβλημα το οποίο έχουμε μοντελοποιήσει, και
- Κατά πόσο η προσέγγισή μας -που βασίζεται στην επίλυση του προβλήματος βελτιστοποίησης που διατυπώθηκε στο Κεφάλαιο 3- συνάδει πραγματικά με την βελτίωση επιθυμητών μετρικών σε ένα οικοσύστημα ομόσπονδης μάθησης (όπως απαιτούμενη ενέργεια και ακρίβεια του νευρωνικού δικτύου)

Σε αυτό το κεφάλαιο θα περιγραφούν οι συνθήκες υπό τις οποίες τελέστηκαν τα πειράματα. Θα αναφερθούν οι επιλογές των παραμέτρων που καθορίζουν τα υπολογιστικά χαρακτηριστικά των συσκευών, το προσομοιωμένο δίκτυο ομόσπονδης μηχανικής μάθησης, ενώ θα περιγραφεί και η διεργασία μηχανικής μάθησης που καλούμε το σύστημα να διαχειριστεί.

5.1 Παράμετροι του Συστήματος

Το οικοσύστημα ομόσπονδης μηχανικής μάθησης αποτελείται από τις συσκευές που συμμετέχουν στην ΟΜ και το κοινό κανάλι επικοινωνίας, δηλαδή το ασύρματο δίκτυο. Για να το μοντελοποιήσουμε στα πειράματά μας, χρησιμοποιούμε τις παραμέτρους που αναφέρθηκαν στο κεφάλαιο 3, και απαιτούνται για την διατύπωση των μεγεθών που μας ενδιαφέρουν.

Για τα πειράματά μας θα θεωρήσουμε ένα δίκτυο που αποτελείται από 10 συσκευές, με 4 επεξεργαστές σε καθεμία. Οι παράμετροι που αφορούν τα χαρακτηριστικά υπολογισμού των συσκευών, το ασύρματο δίκτυο NOMA, και τις QoS απαιτήσεις για τον χρόνο υπολογισμού και επικοινωνίας, δίνονται στον πίνακα 5.1, και ισχύουν για όλα τα πειράματα, αν δεν αναφέρεται κάτι διαφορετικό.

Σταθερά	Εξήγηση	Τιμή
N	Πλήθος συσκευών στο δίκτυο ΟΜ	10
Q_n	Αριθμός επεξεργαστών σε κάθε συσκευή	4
f_{min}, f_{max}	Όρια συχνότητας λειτουργίας επεξεργαστών	$1,3$ GHz
C_n^q	Συντελεστές ενεργ. κατανάλωσης επεξεργαστών (3.20)	$\sim U(0.01, 0.1)$ [1]
P_{const}	Ισχύς αδρανούς λειτουργίας επεξεργαστή	500 mW
p_{max}	Μέγιστη ισχύς μετάδοσης	1 W
B	Εύρος ζώνης του συστήματος	20 MHz
ρ	Απώλεια διαδρομής σε 1m	-15.3 dB
α	Εκθέτης απώλειας διαδρομής	3.74
N_0	Φασματική πυκνότητα ισχύος AWGN θορύβου	-134 dBm/Hz
T_{max}	Χρονικό όριο γύρου επικοινωνίας	0.5 sec
T_{max}^{cmp}	Χρονικό όριο διεργασίας υπολογισμού	$0.8 \cdot T_{max}$
T_{max}^{tx}	Χρονικό όριο μετάδοσης της παραγώγου	$0.8 \cdot T_{max}$

Πίνακας 5.1: Τιμές των σταθερών παραμέτρων του συστήματος

5.2 Διεργασία Μηχανικής Μάθησης

Για τον κώδικα της λειτουργίας του συστήματος ομόσπονδης μηχανικής μάθησης χρησιμοποιήθηκε ως βάση το αποθετήριο του [43], που χρησιμοποιεί την βιβλιοθήκη μηχανικής μάθησης Tensorflow μαζί με κλάσεις που αναπαριστούν τους πελάτες και τον εξυπηρετητή στο σύστημα ομόσπονδης μηχανικής μάθησης. Σε αυτόν τον κώδικα χρειάστηκε να ενσωματωθεί η λειτουργία της επιλογής των δεδομένων με βάση τη σημαντικότητα.

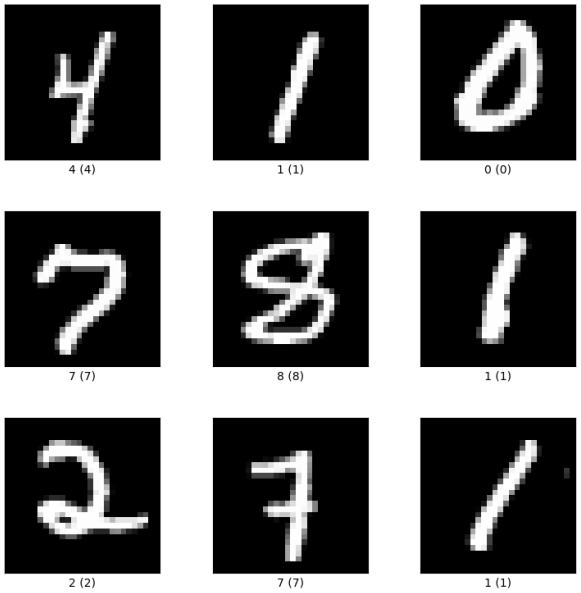
5.2.1 Το Σύνολο Δεδομένων MNIST

Ακολουθώντας τη σχετική με το θέμα της εργασίας διεθνή βιβλιογραφία, η διεργασία μηχανικής μάθησης που επιλέχθηκε για την εκπαίδευση μέσω του συστήματος ομόσπονδης μάθησης είναι η ταξινόμηση σε κλάσεις πάνω στο σύνολο δεδομένων MNIST.

Το MNIST πρόκειται για συλλογή χειρόγραφων ψηφίων. Τα ψηφία αυτά παρέχονται σε μορφή ασπρόμαυρων εικόνων, μεγέθους 28x28 pixels. Αποτελείται από ένα σύνολο εκπαίδευσης 60.000 εικόνων και ένα σύνολο ελέγχου 10.000 εικόνων. Αποτελεί ένα από τα πλέον διαδεδομένα datasets λόγω της ευκολίας χρήσης- χρειάζεται ελάχιστη προσπάθεια για την προεπεξεργασία των δεδομένων [23].

Στο προσομοιωμένο σύστημα ομόσπονδης μάθησης, οι 60.000 εικόνες του συνόλου εκπαίδευσης είναι ισόποσα και τυχαία διαμοιρασμένες στους 10 πελάτες. Δηλαδή, θα εστιάσουμε στην περίπτωση όπου τα δεδομένα είναι κατανεμημένα με iid τρόπο, εφόσον στο πλαίσιο αυτής της εργασίας δεν ασχολούμαστε με τα προβλήματα που εγείρει η non-iid κατανομή των δεδομένων στους πελάτες της ομόσπονδης μάθησης. Το σύνολο ελέγχου χρησιμοποιείται για

[1] Επιλέγονται ομοιόμορφα από το διάστημα των 0.01 έως 0.1 Watt(GFLOPs/s)$^{-3}$



Σχήμα 5.1: Εικόνες από το σύνολο MNIST [8]

την αξιολόγηση του συναθροισμένου μοντέλου μετά από κάθε γύρο επικοινωνίας.

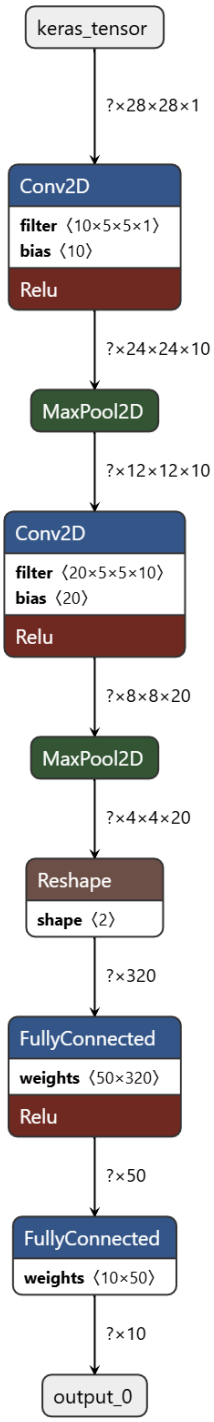
5.2.2 Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου (CNN)

Το μοντέλο που καλείται το σύστημα ομόσπονδης μάθησης να εκπαιδεύσει είναι ένα συνελικτικό νευρωνικό δίκτυο. Αυτό αποτελείται από δύο συνελικτικά επίπεδα με το καθένα να ακολουθείται από συνάρτηση ενεργοποίησης ReLU, και ένα επίπεδο max pooling για την εξαγωγή της μέγιστης τιμής ανά χωρίο διαστάσεων 2x2. Στο τέλος υπάρχει ένα πλήρως συνδεδεμένο επίπεδο με 50 νευρώνες που λαμβάνει μια flattened είσοδο και την τροφοδοτεί σε πλήρως συνδεδεμένο επίπεδο με 10 νευρώνες, από το οποίο προκύπτει η έξοδος με τις προβλέψεις για τις κλάσεις των ψηφίων 0-9. Στο επίπεδο εξόδου χρησιμοποιείται η συνάρτηση ενεργοποίησης Softmax και για συνάρτηση απώλειας η Categorical Cross Entropy.

Οι παράμετροι του μοντέλου είναι 21.840, ενώ τα FLOPs που απαιτούνται για τον χειρισμό ενός δείγματος υπολογίζονται στα λίγο παραπάνω από 1.000.000 και το μέγεθος των παραμέτρων του μοντέλου ανέρχεται στα 723.04 Kbits. Η αρχιτεκτονική του CNN μοντέλου δίνεται και αναλυτικά σε μορφή διαγράμματος στο σχήμα 5.2, που παράχθηκε με χρήση της εφαρμογής Netron [34, 35]. Τέλος, ο ρυθμός μάθησης η επιλέγεται στα 10^{-3}.

5.3 Τοπολογία Ασύρματου Δικτύου

Για να υπολογιστεί η χρονική καθυστέρηση (και κατ' επέκταση η απαιτούμενη ενέργεια) για την επικοινωνία των συσκευών με τον εξυπηρετητή-σταθμό βάσης το τηλεπικοινωνιακό



Σχήμα 5.2: Διάγραμμα της αρχιτεκτονικής του CNN που παράχθηκε μέσω της εφαρμογής Netron

πρωτόκολλο NOMA χρησιμοποιεί τα κέρδη καναλιού, που εξαρτώνται αποκλειστικά από την απόσταση κάθε χρήστη από τον σταθμό βάσης. Οι συσκευές τοποθετούνται τυχαία στο χώρο σε μια ακτίνα 200m από τον εξυπηρετητή, και η απώλεια διαδρομής PL_n υπολογίζεται σύμφωνα με τον τύπο 5.1 για συνθήκες σε αστικές περιοχές [9] και αντιστοιχεί στις παραμέτρους ρ, α που αναφέρονται στον πίνακα 5.1.

$$PL_n = 128.1 + 37.6 \log_{10}(\frac{d_n}{1000}) \text{ [dB]}, \tag{5.1}$$

όπου d_n είναι η απόσταση της συσκευής από τον εξυπηρετητή σε μέτρα.

Η απώλεια διαδρομής για μια συσκευή n είναι αντίστροφη του κέρδους του καναλιού της, δηλ. ισχύει ότι

$$G_n = \frac{1}{PL_n} \tag{5.2}$$

5.4 Κατανομή των Δεδομένων Εκπαίδευσης στους Συμμεντέχοντες

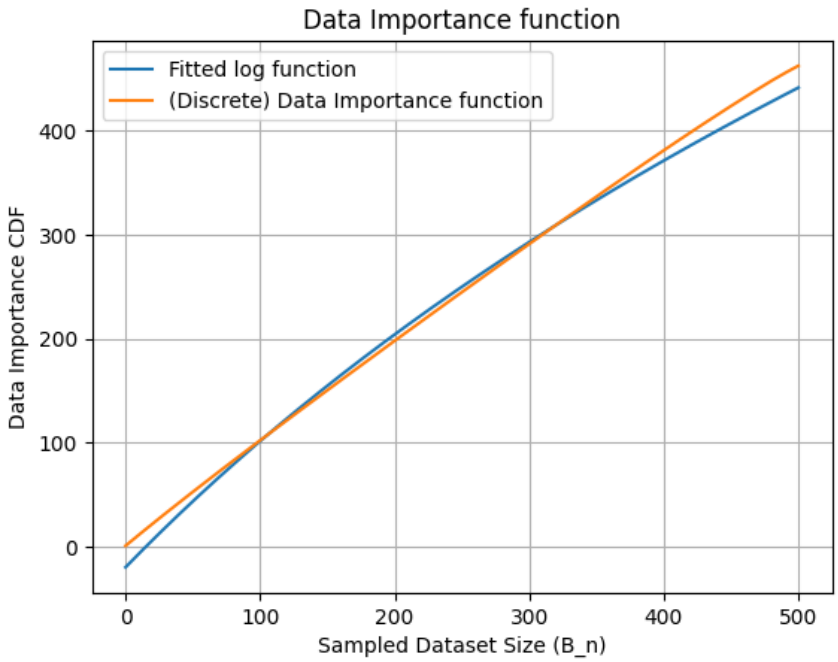
Στα περισσότερα πειράματα τα δεδομένα είναι τυχαία και ισόποσα διαμοιρασμένα στις συσκευές του συστήματος, οπότε είναι ανεξάρτητα και ομοιόμορφα κατανεμημένα (Independently and Identically Distributed - IID). Ωστόσο, θα μελετηθεί σε κάποια πειράματα και το σενάριο όπου τα δεδομένα είναι σε μορφή βασισμένη στην pathological non-IID [30], με τους χρήστες να έχουν μόνο δεδομένα από 1-2 ετικέτες από τις 10 συνολικές.

5.5 Σημεία Αναφοράς για την Αξιολόγηση της Λύσης (Benchmarks)

Προκειμένου να εκτιμήσουμε την επιτυχία της προτεινόμενης λύσης στην επίλυση του διατυπωμένου προβλήματος και στην βελτίωση των καίριων μετρικών που αφορούν την ομόσπονδη μάθηση, θα πραγματοποιήσουμε μια συγκριτική αξιολόγηση αντιπαραβάλλοντας τη λύση μας με εναλλακτικές, πιο συμβατικές προσεγγίσεις στο πρόβλημα. Αυτές οι προσεγγίσεις, που θα δοκιμάσουμε να αντικαταστήσουν είτε συστατικά στοιχεία του προτεινόμενου αλγορίθμου επίλυσης είτε ολόκληρο τον αλγόριθμο, θα περιγραφούν παρακάτω.

5.5.1 Επίλυση του προβλήματος με Αλγόριθμο για Μη Κυρτή Βελτιστοποίηση

Σαν πρώτο σημείο σύγκρισης θα θεωρήσουμε την αριθμητική επίλυση του προβλήματος 3.26 για την εύρεση μιας λύσης για B_n, B_n^q, f_n^q, p_n. Γι'αυτόν τον σκοπό θα χρησιμοποιήσουμε την συνάρτηση minimize της βιβλιοθήκης SciPy με χρήση του επιλυτή 'trust-const', που χρησιμοποιεί περιοχές εμπιστοσύνης (trust regions) για την βελτιστοποίηση. Αυτό σε πολύ γενικές γραμμές σημαίνει ότι η αντικειμενική συνάρτηση επαναληπτικά προσεγγίζεται (συνήθως με τετραγωνικούς όρους) μέσα σε μια περιοχή γύρω από ένα τρέχον σημείο, και το εύρος αυτής της περιοχής προσαρμόζεται ανάλογα με την ακρίβεια αυτής της προσέγγισης [7, 38].



Σχήμα 5.3: Προσέγγιση της συνάρτησης σημαντικότητας $g_n(B_n)$ με λογαριθμική συνάρτηση για μια συσκευή.

Η συνάρτηση $g_n(B_n)$ που ορίστηκε κατά την μοντελοποίηση του συστήματος (εξ. 3.12) αν και είναι χρήσιμη για την δική μας προσέγγιση του προβλήματος, σε μια τέτοια μέθοδο αριθμητικής επίλυσης θα ήταν καλύτερη η χρήση μιας συνεχούς συνάρτησης από την επέκταση της διακριτής $g_n(B_n)$. Γι'αυτό, θα αξιοποιήσουμε το γεγονός ότι η $g_n(B_n)$ είναι κοίλη (βλ. Λήμμα 3.2) για να βρούμε μια λογαριθμική συνάρτηση της μορφής $a\log(1+bx)$ που ταιριάζει στα σημεία της $g_n(B_n)$, ελαχιστοποιώντας το τετραγωνικό σφάλμα. Γίνεται λοιπόν χρήση της συνάρτησης curve_fit της SciPy. Το αποτέλεσμα για τη συνάρτηση σημαντικότητας μίας συσκευής φαίνεται στο διάγραμμα του σχήματος 5.3.

5.5.2 Εναλλακτική Στρατηγική Επιλογής Δεδομένων

Στη συνέχεια θα αντιπαραβάλουμε τη λύση μας με τη χρήση διαφορετικών στρατηγικών για την επιλογή των δεδομένων σε κάθε συσκευή. Η επιλογή θα γίνεται και πάλι με βάση τη σημαντικότητα όπως αναλύθηκε στην υποενότητα 3.2. Ο αλγόριθμος παραμένει ίδιος με τον προτεινόμενο ως προς την ανάθεση πόρων, όμως, τα δείγματα σε κάθε συσκευή επιλέγονται ντετερμινιστικά ως ακολούθως:

- Επιλογή ολόκληρου του προσωπικού συνόλου δεδομένων
- Επιλογή ποσοστού του προσωπικού συνόλου δεδομένων (π.χ. 10%, 25%, 50%)

5.5.3 Εναλλακτική Στρατηγική Ανάθεσης Ραδιοπόρων και Υπολογιστικών Πόρων

Αντίστοιχα, για την αξιολόγηση της στρατηγικής ανάθεσης πόρων που χρησιμοποιούμε, δημιουργούμε τα κάτωθι σενάρια όπου αντικαθιστούμε συστατικά στοιχεία του προτεινόμενου αλγορίθμου επίλυσης με μια πιο απλοϊκή στρατηγική για την ανάθεση συχνοτήτων λειτουργίας επεξεργαστών ή ισχύων μετάδοσης. Η στρατηγική επιλογής δεδομένων δεν αλλάζει.

- Μετάδοση από όλες τις συσκευές στη μέγιστη δυνατή ισχύ, δηλ. στο 1 W.
- Μετάδοση από όλες τις συσκευές στη μισή από τη μέγιστη δυνατή ισχύ, δηλ. στα 0.5 W.
- Σταθερή συχνότητα επεξεργασίας στα 2GHz από όλους τους επεξεργαστές (στη μέση του διαστήματος $[f_{min}, f_{max}]$)

6

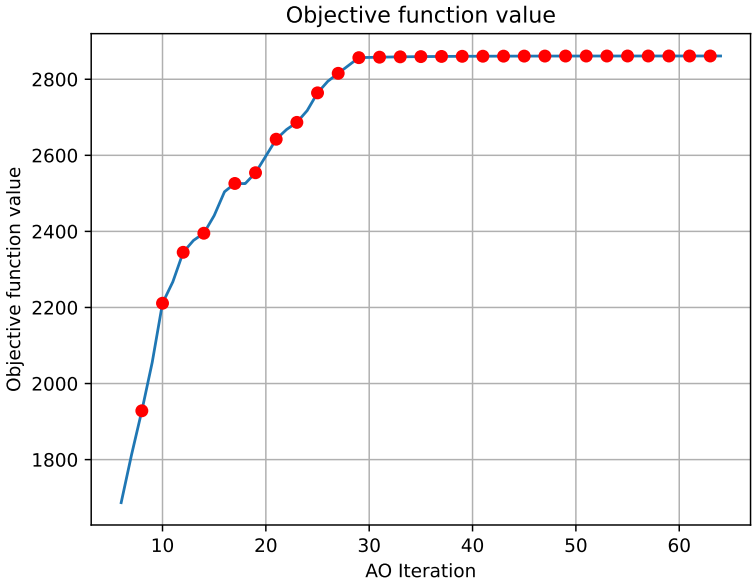
Αποτελέσματα

Σε αυτό το κεφάλαιο παρουσιάζονται τα αποτελέσματα των πειραμάτων όπως αυτά περιγράφηκαν στο προηγούμενο κεφάλαιο. Αρχικά εξετάζεται ο τρόπος με τον οποίον συγκλίνει ο αλγόριθμος βελτιστοποίησης. Έπειτα, αξιολογούνται μετρικές όπως η ακρίβεια του συνολικού μοντέλου και η ενέργεια που καταναλώνεται, σε σύγκριση και με τα benchmarks που ορίζουμε στην υποενότητα 5.5. Αυτό γίνεται με την παρακολούθηση των μετρικών είτε κατά τη διάρκεια των γύρων επικοινωνίας της ομόσπονδης μάθησης είτε κατά την κλιμάκωση του δικτύου με την είσοδο περισσότερων συσκευών.

Υπενθυμίζεται ότι οι παράμετροι του συστήματος είναι αυτοί που αναγράφονται στον πίνακα 5.1, εκτός αν αναφέρεται κάτι διαφορετικό για κάποιο πείραμα. Επίσης, για τις μετρικές που αφορούν τη γενική συμπεριφορά του συστήματος, τα αποτελέσματα που παρουσιάζονται είναι ο μέσος όρος που έχει προκύψει από προσομοιώσεις με διαφορετικές τυχαίες αρχικοποιήσεις για την τοποθέτηση των συσκευών στο χώρο, για την κατανομή των προσωπικών συνόλων εκπαίδευσης, και για τους σταθερούς συντελεστές ενεργειακής κατανάλωσης των επεξεργαστών C_n^q.

6.1 Σύγκλιση της Αναλυτικής Λύσης

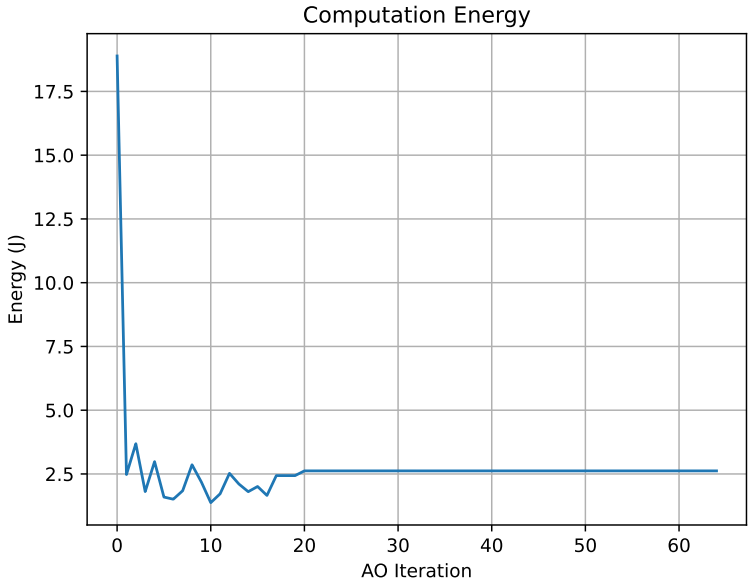
Αρχικά, θα παρατηρήσουμε την πορεία της βελτιστοποίησης σύμφωνα με τους αλγόριθμους 2 και 3 για έναν γύρο επικοινωνίας. Το διάγραμμα του σχήματος 6.1, αναδεικνύει την πορεία που ακολουθεί η παράμετρος Dinkelbach y, και άρα η αντικειμενική συνάρτηση, κατά την εκτέλεση του αλγορίθμου βελτιστοποίησης για έναν γύρο επικοινωνίας. Είναι φανερή η απότομη αύξηση της τιμής της από τις αρχικές επαναλήψεις της εναλλασσόμενης βελτιστοποίησης. Η στασιμότητα στην τιμή έπειτα σημαίνει ότι ο αλγόριθμος έχει συγκλίνει και πρακτικά θα έληγε σε νωρίτερη επανάληψη, ωστόσο για λόγους πληρότητας εν προκειμένω η εκτέλεση έγινε για μεγαλύτερο αριθμό επαναλήψεων. Στο διάγραμμα είναι επισημειωμένα τα σημεία στα οποία ανανεώνεται η παράμετρος y του αλγορίθμου Dinkelbach. Αυτό έχει οριστεί να γίνεται όταν σε έναν γύρο ανανέωσης των $\{B_n^q\}, \{f_n^q\}, \{p_n\}$ η ανανεωμένη τιμή του y διαφέρει ελάχιστα από την προηγούμενη, δηλαδή $(y[t+1]-y[t])/y[t] < 1\%$. Εναλλακτικά, αυτό συμβαίνει με το πέρας του μέγιστου αριθμού επαναλήψεων.



Σχήμα 6.1: Σύγκλιση της τιμής της αντικειμενικής συνάρτησης. Με κόκκινες τελείες είναι επισημειωμένες οι ανανεώσεις στην παράμετρο y του αλγορίθμου Dinkelbach.

Η ενέργεια υπολογισμού και επικοινωνίας στα σχήματα 6.2, 6.3 εμφανίζουν πτωτική πορεία κατά τη διάρκεια της βελτιστοποίησης, κάτι που είναι αναμενόμενο εφόσον σημαντικό κομμάτι του στόχου της συνιστά η ελάττωση της καταναλισκόμενης ενέργειας για τον γύρο επικοινωνίας. Στο σχήμα 6.2 παρατηρείται στην αρχή μια κατακόρυφη πτώση στην ενέργεια υπολογισμού λόγω της απομάκρυνσης από την αρχική εφικτή λύση ("μαντεψιά"). Έπειτα, η τιμή αυξομειώνεται καθώς μέχρι να συγκλίνει, επηρεάζεται από τις αλλαγές κατά την επίλυση των υποπροβλημάτων τόσο της επιλογής δεδομένων $\mathcal{P}2$ (4.3) όσο και της ανάθεσης υπολογιστικών πόρων $\mathcal{P}3$ (4.10)- Μπορεί αμφότερα να στοχεύουν στην μεγιστοποίηση της αντικειμενικής συνάρτησης, όμως το υποπρόβλημα της ανάθεσης δεδομένων μπορεί να αυξήσει την ενέργεια υπολογισμού με την ανάθεση περισσότερων δεδομένων, και η νέα λύση για B_n^q μπορεί να επηρεάσει τη λύση του προβλήματος ανάθεσης συχνοτήτων αλλάζοντας τον χρονικό περιορισμό και οδηγώντας ενδεχομένως στην αύξηση κάποιων συχνοτήτων λειτουργίας.

Από την άλλη, η γενικά πτωτική πορεία της ενέργειας για την μετάδοση στο σχήμα 6.3 είναι ομαλότερη, κάτι που είναι σχετικά αναμενόμενο εφόσον το πρόβλημα 4.16 δεν περιέχει στην αντικειμενική του συνάρτηση μεταβλητές που διαμοιράζεται με τα προβλήματα $\mathcal{P}2$ και $\mathcal{P}3$. Αυτό βέβαια δεν σημαίνει ότι το πρόβλημα είναι εντελώς ανεξάρτητο από τα υπόλοιπα, εφόσον ο χρόνος μετάδοσης συμπεριλαμβάνεται στους χρονικούς περιορισμούς όλων των υποπροβλημάτων, και αντίστοιχα για τον χρόνο υπολογισμού στο πρόβλημα του power control. Εκεί ενδέχεται να οφείλεται και η ελαφριά αύξηση στην ενέργεια κατά τις πρώτες επαναλήψεις της εναλλασσόμενης βελτιστοποίησης, που γίνεται παράλληλα με την μείωση της ενέργειας υπολογισμού, πιθανώς μέσω της μείωσης των συχνοτήτων λειτουργίας, και άρα της αύξησης του χρόνου υπολογισμού που θα ωθούσε τις συσκευές να αυξήσουν την ισχύ μετάδοσης για να

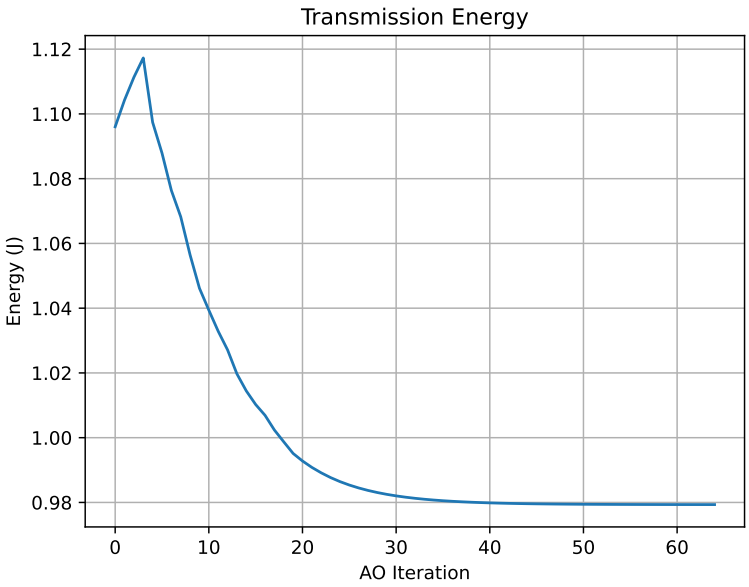


Σχήμα 6.2: Σύγκλιση της ενέργειας υπολογισμού

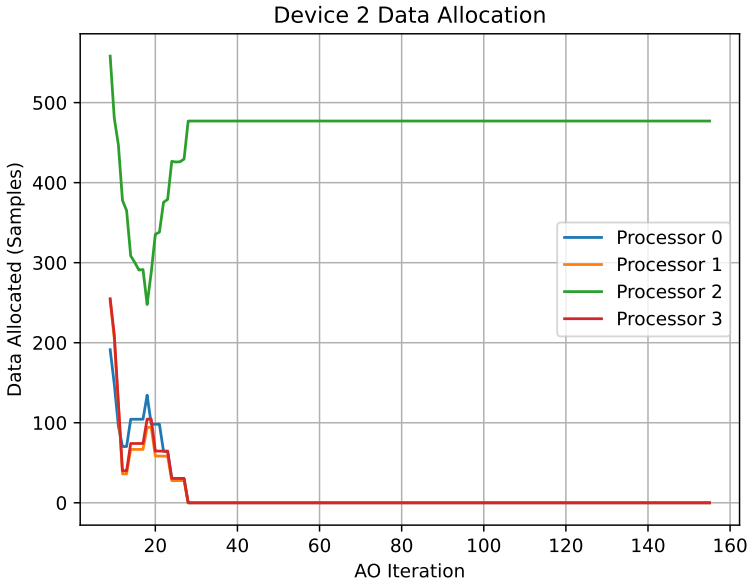
μεταδώσουν την ανανέωση γρηγορότερα.

Στο διάγραμμα 6.4 φαίνεται η διαμόρφωση της ανάθεσης των δεδομένων στους επεξεργαστές μιας συσκευής κατά τη διάρκεια της βελτιστοποίησης. Στην αρχή παρατηρείται η μείωση των δειγμάτων σε όλους τους επεξεργαστές, που δείχνει την υπερίσχυση του στόχου της μείωσης της ενέργειας υπολογισμού έναντι της αύξησης της σημαντικότητας των επιλεγμένων δεδομένων. Ωστόσο, μετά από ένα σημείο σταματούν να ελαττώνονται τα συνολικά δεδομένα σε όλους τους επεξεργαστές, και φαίνεται ότι περισσότερη έμφαση δίνεται στην διαφορετική ανάθεση των δεδομένων στους επεξεργαστές, προκειμένου οι πιο ενεργειακά αποδοτικοί επεξεργαστές να αναλάβουν περισσότερα δεδομένα προς την συνολική μείωση της ενέργειας. Στην παρούσα διαδικασία βελτιστοποίησης μάλιστα, καταλήγουν οι 3 από τους 4 επεξεργαστές της συσκευής να παραμένουν κλειστοί για τον γύρο επικοινωνίας.

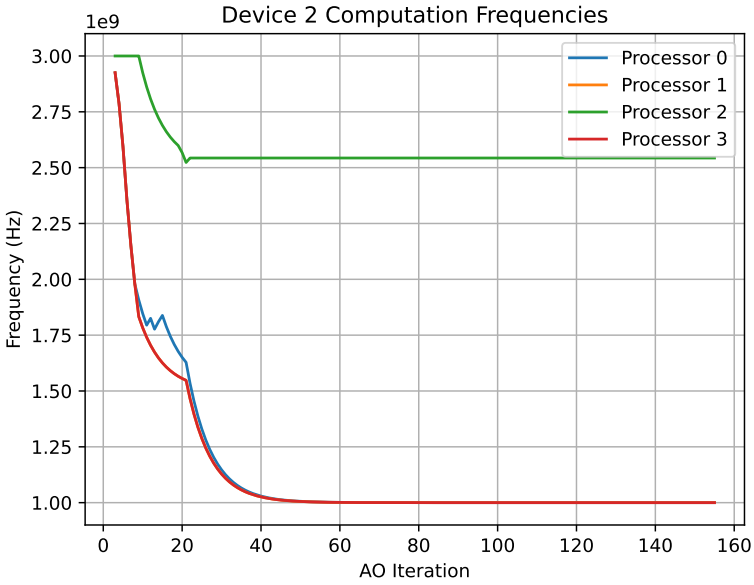
Το παραπάνω διάγραμμα αξίζει να εξεταστεί παράλληλα με το διάγραμμα 6.5, καθώς παρατηρείται η παρόμοια πτωτική πορεία στην αρχή για όλους τους επεξεργαστές, και για τον επεξεργαστή 2 παρατηρείται η σύγκλιση σε μία συχνότητα την ίδια στιγμή που άρχισαν να αυξάνονται τα δείγματα που του αναθέτονται. Μπορεί να γίνει με σχετική ασφάλεια η υπόθεση ότι αυτή η συχνότητα πρόκειται για την λύση $f_n^{q,Energy}$ στην εξίσωση 4.14, η οποία εξαρτάται μόνο από τα χαρακτηριστικά του επεξεργαστή και την ισχύ της αδρανούς λειτουργίας του P_{const}. Από εκεί και έπειτα, συνεχίζει σύμφωνα με αυτήν την υπόθεση να βελτιστοποιείται μόνο η ανάθεση των δειγμάτων, εφόσον γίνεται φανερό ότι δεν επηρεάζεται σε αυτό το σημείο από τον χρονικό περιορισμό. Παρατηρείται επίσης καθ'όλη τη διάρκεια της βελτιστοποίησης η ίδια ιεράρχηση στην ανάθεση του πλήθους δειγμάτων και των συχνοτήτων επεξεργασίας. Μπορούμε να συμπεράνουμε ότι αυτό σχετίζεται άμεσα με την ενεργειακή αποδοτικότητα του κάθε επεξεργαστή - δηλ. ότι ανατίθενται περισσότερα δείγματα στον πιο αποδοτικό επεξεργα-



Σχήμα 6.3: Σύγκλιση της ενέργειας για την μετάδοση



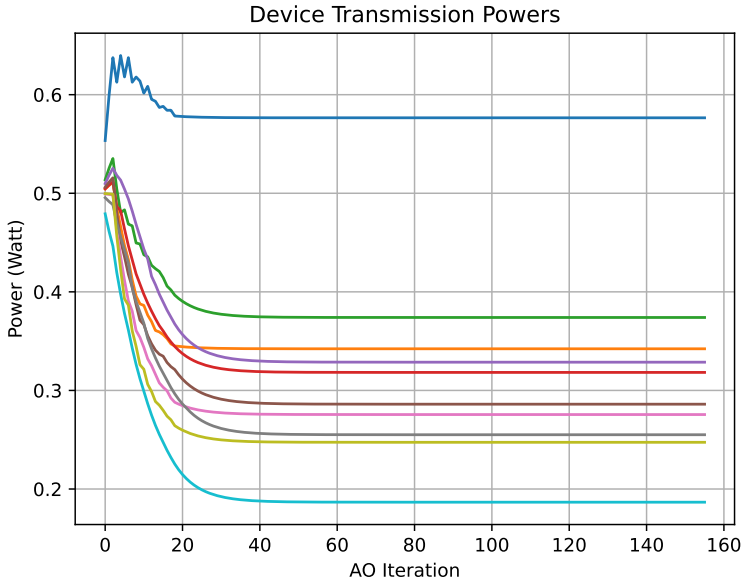
Σχήμα 6.4: Σύγκλιση της ανάθεσης δεδομένων στους επεξεργαστές της συσκευής 2



Σχήμα 6.5: Σύγκλιση στις συχνότητες λειτουργίας των επεξεργαστών της συσκευής 2

στή της συσκευής και ως αποτέλεσμα είναι μεγαλύτερη και η συχνότητα λειτουργίας για να ικανοποιηθεί ο χρονικός περιορισμός (αναδεικνύοντας ότι σε αυτήν την περίπτωση προτιμάται το trade-off της ενέργειας που απαιτεί ο χειρισμός παραπάνω δειγμάτων).

Αντίστοιχα, στη σύγκλιση των ισχύων μετάδοσης μπορεί να παρατηρηθεί μια διάταξη που είναι συνδεδεμένη με την διάταξη των κερδών των συσκευών στο κανάλι, όμως σίγουρα εξαρτάται επίσης από τον χρόνο που απομένει σε κάθε συσκευή για την μετάδοση με δεδομένο τον αντίστοιχο χρόνο υπολογισμού της. Όπως φαίνεται εξάλλου από τις εξισώσεις 4.18, η ισχύς μετάδοσης μιας συσκευής επηρεάζεται κυρίως από τις συνθήκες του καναλιού (εύρος ζώνης, κέρδος συσκευής), από τον εναπομείναντα χρόνο για την μετάδοση της συσκευής, και από τις παρεμβολές των συσκευών με μικρότερο κέρδος καναλιού.



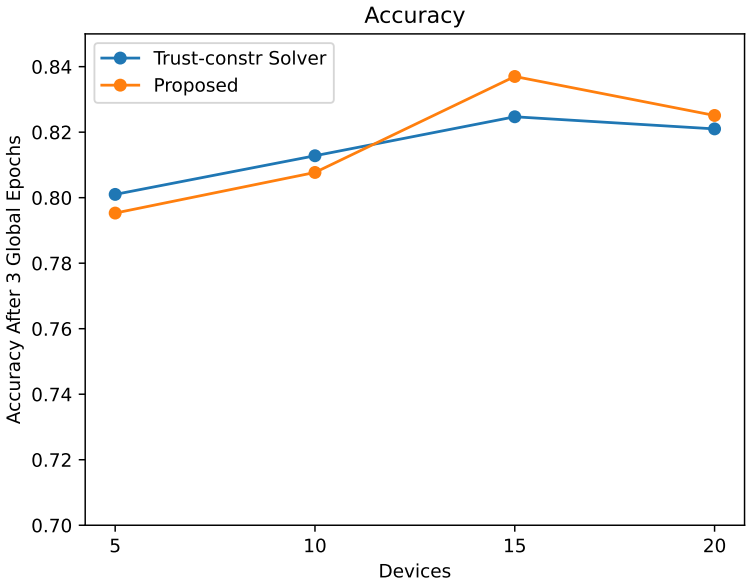
Σχήμα 6.6: Σύγκλιση στις ισχύς μετάδοσης των συσκευών

6.2 Αξιολόγηση της Αναλυτικής Λύσης

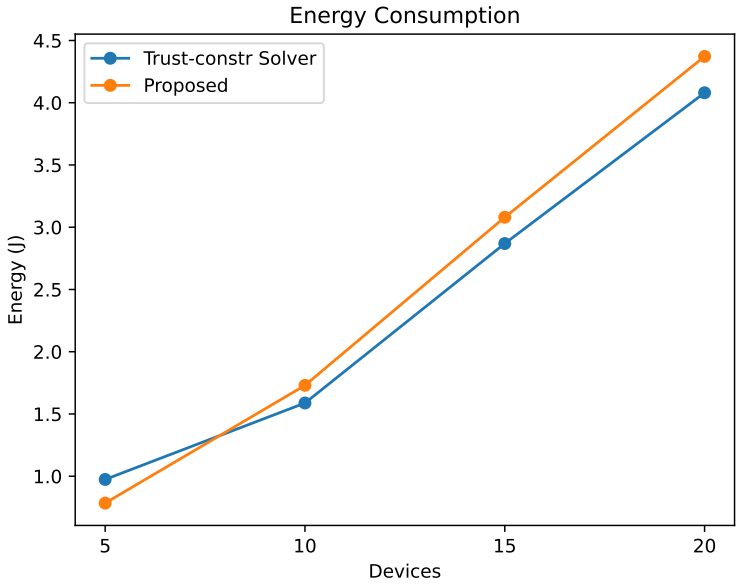
Εδώ θα παρουσιάσουμε τα αποτελέσματα της σύγκρισης του προτεινόμενου αλγορίθμου με την επίλυση με τη χρήση του αλγορίμου trust-constr για την επίλυση μη κυρτών προβλημάτων, όπως αναφέρθηκε στην υποενότητα 5.5.1. Επειδή θα εξετάσουμε τους αλγορίθμους κατά την κλιμακωσιμότητα του δικτύου ΟΜ, το χρονικό όριο T_{max} θα αυξηθεί σε 1 sec, για να μπορεί να υποστηριχθεί η ταυτόχρονη μετάδοση από 20 συσκευές στο σύστημα NOMA, διατηρώντας έναν ρυθμό μετάδοσης μεγαλύτερο από 1 Mbps για κάθε συσκευή. Επίσης σε κάθε συσκευή θα περιέχονται σταθερά 1.000 δείγματα, προκειμένου με την εισαγωγή νέων συσκευών στο σύστημα ΟΜ να αυξάνεται η "γνώση" που περιέχεται σε αυτό.

Όπως φαίνεται από τα σχήματα 6.7, 6.8, ο αλγόριθμος trust-constr αποφέρει αρκετά καλά αποτελέσματα, κρίνοντας από την ακρίβεια του μοντέλου στο σύνολο ελέγχου μετά από 4 γύρους επικοινωνίας και από την μέση ενέργεια που καταναλώνεται σε αυτούς τους γύρους.

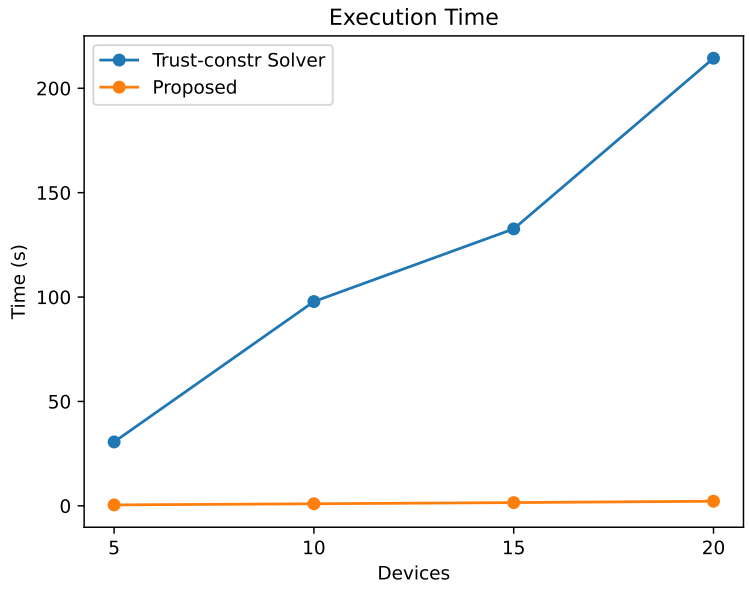
Εφόσον τα αποτελέσματα είναι τόσο κοντινά σε αυτά του προτεινόμενου αλγορίθμου, μπορεί να εγερθεί η απορία για τον λόγο μη προτίμησης αυτού του αλγορίθμου γενικής χρήσης έναντι του προτεινόμενου. Η απάντηση έγκειται στην εξέταση που έγινε για τον χρόνο εκτέλεσης του κάθε αλγορίθμου, και παρουσιάζεται στα διαγράμματα των σχημάτων 6.9, 6.10. Ο επιλυτής που χρησιμοποιείται για προβλήματα περιορισμένης μη κυρτής βελτιστοποίησης παρουσιάζει απαγορευτικό χρόνο εκτέλεσης για την βελτιστοποίηση των παραμέτρων σε έναν γύρο επικοινωνίας. Αυτό αναδεικνύει την ανάγκη της αξιοποίησης των χαρακτηριστικών του προβλήματος για την χρήση ενός εξειδικευμένου αλγορίθμου που φτάνει αποδοτικά σε μια ικανοποιητική λύση, ενώ η ύπαρξη πολλών μεταβλητών απόφασης κατά τη κλιμάκωση του δικτύου αναδεικνύει αντίστοιχα τα οφέλη της διάσπασης του προβλήματος σε υποπροβλήματα με αισθητά λιγότερες μεταβλητές. Παρατηρώντας την πορεία του χρόνου εκτέλεσης του προτεινόμενου αλγόριθμου στο διάγραμμα 6.10 φαίνεται ότι ο χρόνος αυξάνεται γραμμικά κατά την κλιμάκωση του δικτύου ώστε να περιέχει από 5 μέχρι 20 συσκευές, κάτι το οποίο είναι επιθυμητό.



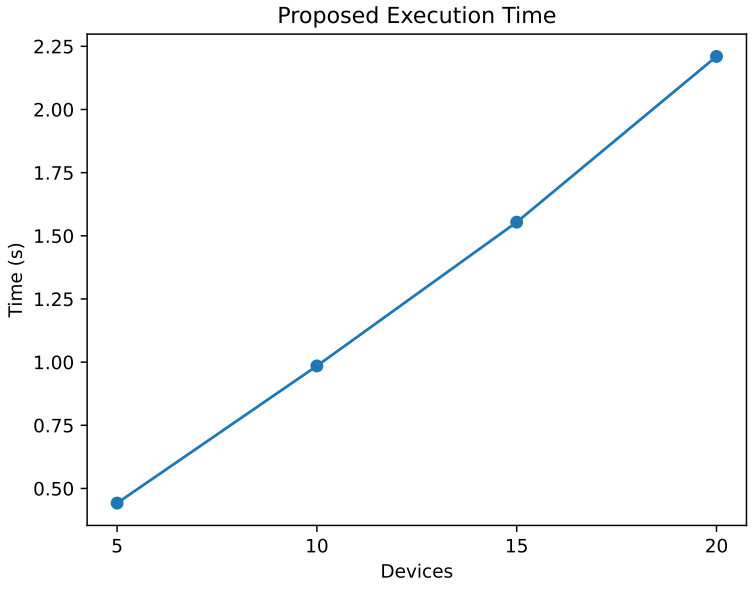
Σχήμα 6.7: Ακρίβεια που επιτυγχάνεται μετά από 4 γύρους επικοινωνίας για αυξανόμενο αριθμό συσκευών στο δίκτυο OM



Σχήμα 6.8: Μέση ενέργεια κατανάλωσης ανά γύρο επικοινωνίας για αυξανόμενο αριθμό συσκευών στο δίκτυο OM



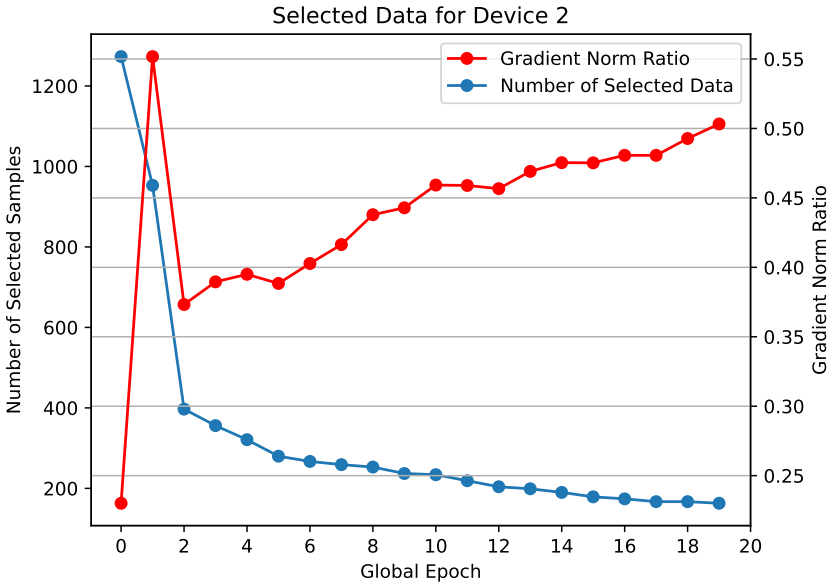
Σχήμα 6.9: Σύγκριση των χρόνων εκτέλεσης των διαφορετικών αλγορίθμων επίλυσης του προβλήματος βελτιστοποίησης



Σχήμα 6.10: Χρόνος εκτέλεσης του προτεινόμενου αλγορίθμου κατά την κλιμάκωση του δικτύου OM

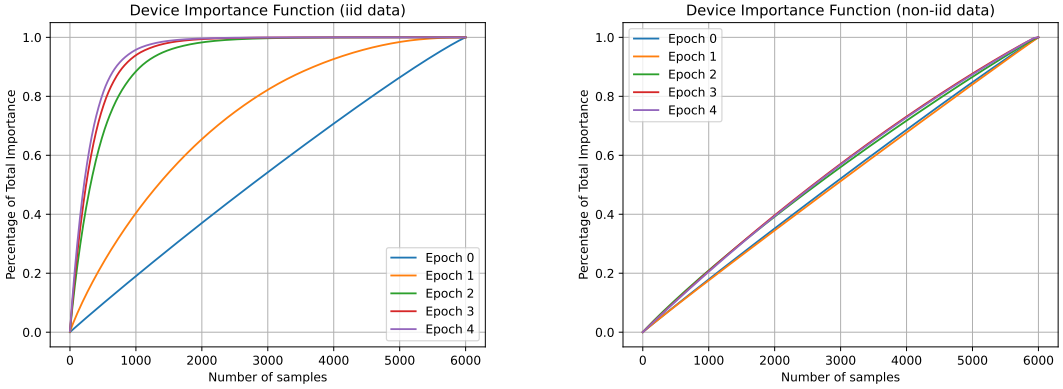
6.3 Αξιολόγηση της Στρατηγικής Επιλογής Δεδομένων

Θα προχωρήσουμε με την αξιολόγηση της στρατηγικής επιλογής δεδομένων που αναλύθηκε στην υποενότητα 3.2, και οδήγησε στην διατύπωση του αλγορίθμου 1. Αρχικά, θα σχολιάσουμε την συμπεριφορά της στρατηγικής δεδομένων κατά την εκπαίδευση ενός μοντέλου κατά την ομόσπονδη μάθηση. Στο σχήμα 6.11 βρίσκεται σε κοινή γραφική παράσταση ο αριθμός των επιλεγμένων δειγμάτων σε μία συσκευή, καθώς και το ποσοστό της εκτιμώμενης νόρμας της παραγώγου που τα δεδομένα αυτά ενέχουν ως ποσοστό του συνόλου. Φαίνεται ότι ο αριθμός των επιλεγμένων δεδομένων μειώνεται δραματικά κατά την εκπαίδευση του μοντέλου, από περίπου 1200 δείγματα, ή το 20% του προσωπικού συνόλου δεδομένων, σε περίπου 150 δείγματα, ή το 2.5% του τοπικού συνόλου. Ταυτόχρονα, τα επιλεγμένα δείγματα, αν και λιγότερα, εν γένει περιέχουν μεγαλύτερο ποσοστό της σημαντικότητας όλων των δεδομένων της συσκευής - ακόμα και το 2.5% των δεδομένων στο τέλος της εκπαίδευσης περιέχουν παραπάνω από το 50% της σημαντικότητας.



Σχήμα 6.11: Αριθμός επιλεγμένων δειγμάτων (αριστερός y άξονας, μπλε χρώμα) και το ποσοστό του μέτρου της παραγώγου που συνιστούν ως επί το ολόκληρο σύνολο δεδομένων (δεξιός y άξονας, κόκκινο χρώμα)

Το παραπάνω διάγραμμα έχει σημασία να το αντιπαραβάλουμε με το διάγραμμα 6.12α΄, που παρουσιάζει πώς διαμορφώνεται η αθροιστική κατανομή της σημαντικότητας των δειγμάτων κατά την διαδικασία της εκπαίδευσης, που ουσιαστικά αποτελεί την συνάρτηση $g_n(B_n)$ που ορίζεται στο 3.12. Αν παρατηρήσουμε την αθροιστική κατανομή της σημαντικότητας των δεδομένων σε μία συσκευή κατά τη διάρκεια της εκπαίδευσης, τότε θα καταστεί εμφανές ότι στην περίπτωση των IID δεδομένων, όσο εκπαιδεύεται το μοντέλο κατά τη διάρκεια της ομόσπονδης μάθησης, η σημαντικότητα των δεδομένων συγκεντρώνεται σε ολοένα και λιγότερα



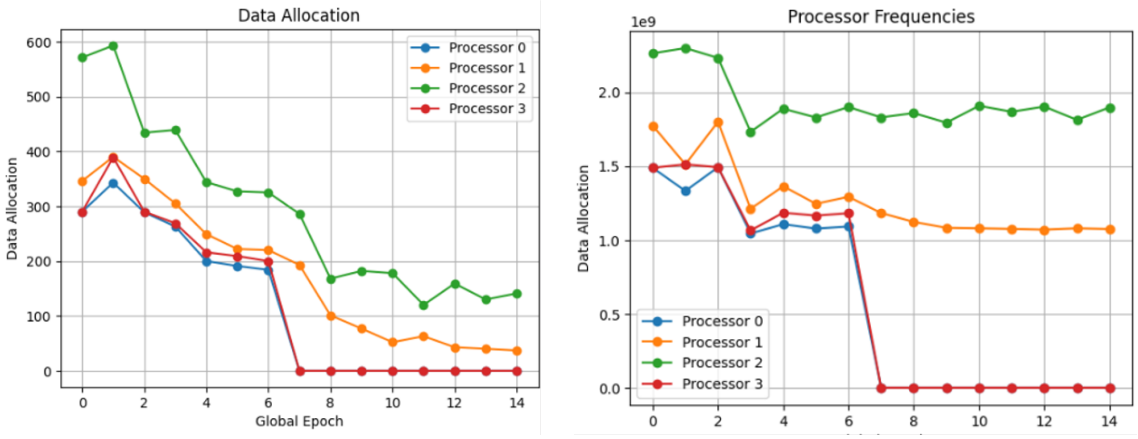
(α΄) IID δεδομένα (β΄) Non-IID δεδομένα

Σχήμα 6.12: Αθροιστική κατανομή της σημαντικότητας των δεδομένων σε μία συσκευή για τους 5 πρώτους γύρους επικοινωνίας του συστήματος ομόσπονδης μάθησης

δεδομένα, εφόσον καταλήγουν να ελαττώνονται τα δείγματα στα οποία το καθολικό μοντέλο ταξινομεί με μεγάλο σφάλμα. Μπορεί συνάμα να εξηγηθεί και η απότομη αύξηση στην γραφική του ποσοστού της παραγώγου στο σχήμα 6.11 στην δεύτερη εποχή, επειδή σε αυτήν έχουν ήδη να ξεχωρίζουν τα δεδομένα με την μεγαλύτερη σημαντικότητα, όμως όχι στον βαθμό που γίνεται στις επόμενες εποχές, με αποτέλεσμα να μην αποθαρρύνεται τόσο η επιλογή περισσότερων δειγμάτων. Ωστόσο, παρατηρώντας το ίδιο για μια παρόμοια διάταξη στο σύστημα με non-iid δεδομένα στις συσκευές, αναδεικνύεται το πρόβλημα που φέρνει αυτή η κατανομή δεδομένων εφόσον πολύ μικρότερη διαφορά παρατηρείται στην κατανομή της σημαντικότητας των δεδομένων. Δηλαδή, σε πολύ μικρότερο βαθμό ξεχωρίζουν τα δεδομένα που δεν ταξινομούνται σωστά, στα οποία πρέπει να επικεντρωθεί η εκπαίδευση.

Στο σχήμα 6.13 φαίνονται σε διπλανές γραφικές παραστάσεις οι πορείες κατά την ΟΜ των δεδομένων που ανατίθενται σε κάθε επεξεργαστή μίας συσκευής, καθώς και οι αντίστοιχες συχνότητες λειτουργίας επεξεργαστών που επιλέγονται από τον προτεινόμενο αλγόριθμο. Όπως είναι αναμενόμενο, όλο και λιγότερα δεδομένα ανατίθενται σε κάθε επεξεργαστή στην πορεία της εκπαίδευσης, ενώ δύο από τους επεξεργαστές κλείνουν μετά από κάποιες εποχές, και δεν αναλαμβάνουν κανένα δείγμα μετά. Παρατηρώντας αυτά τα γραφήματα μπορούμε να αποφανθούμε πάλι για τους επεξεργαστές με την καλύτερη ενεργειακή απόδοση (δηλαδή με τους μικρότερους συντελεστές C_n^q), εφόσον αυτοί "ενθαρρύνονται" από τον αλγόριθμο βελτιστοποίησης να αναλάβουν περισσότερα δείγματα, ενδεχομένως λειτουργώντας σε μεγαλύτερη συχνότητα, προκειμένου οι υπόλοιποι επεξεργαστές να αναλάβουν λιγότερα ή να παραμείνουν ανενεργοί.

Ακολουθεί η σύγκριση του προτεινόμενου αλγορίθμου με παραλλαγές του όπου η στρατηγική ανάθεσης πόρων και η κατανομή δεδομένων στους επεξεργαστές παραμένει ίδιος, όμως στις συσκευές επιλέγεται σταθερός αριθμός δεδομένων. Από τις γραφικές των σχημάτων 6.14 και 6.16 συνάγουμε ότι ο προτεινόμενος αλγόριθμος για την εκπαίδευση στο MNIST φτάνει στα ίδια επίπεδα ακρίβειας μοντέλου με τα σενάρια όπου επιλέγεται αριθμός δεδομένων, ξοδεύοντας λιγότερη ενέργεια, που μάλιστα μειώνεται στη διάρκεια της εκπαίδευσης, επειδή ολοένα



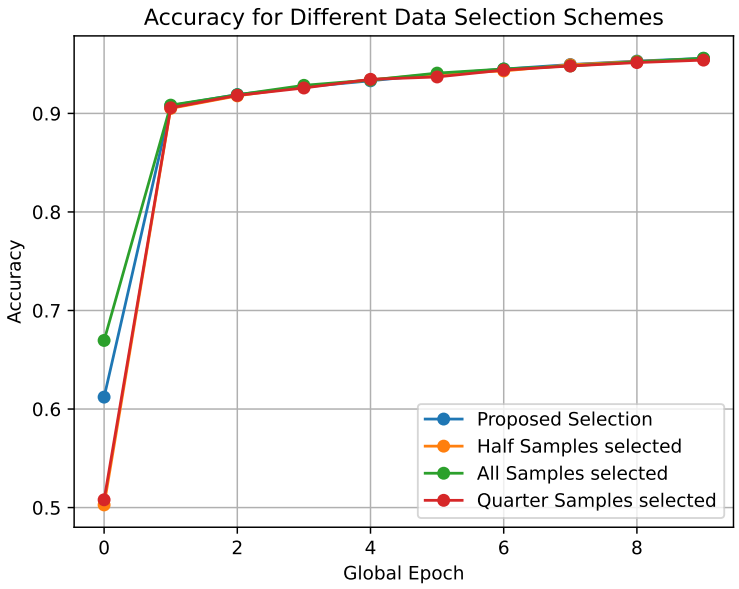
(α΄) Δεδομένα ανά επεξεργαστή (β΄) Συχνότητα λειτουργίας ανά επεξεργαστή

Σχήμα 6.13: Δεδομένα που ανατίθενται και συχνότητες που επιλέγονται για τους επεξεργαστές μίας συσκευής, κατά τη διάρκεια της ΟΜ

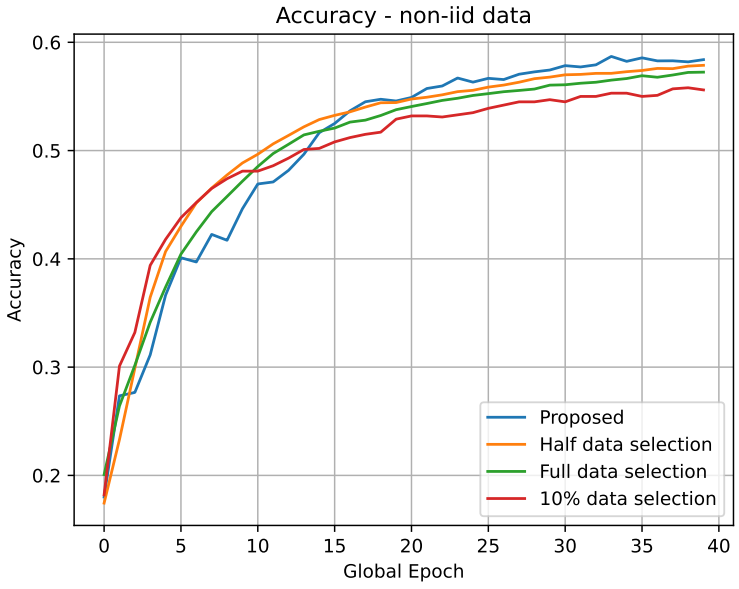
και λιγότερα δείγματα θεωρούνται "σημαντικά" ωστέ να αξίζει η παραπάνω ενέργεια για την εκπαίδευση σε αυτά- εξάλλου τα γραφήματα στα 6.12α΄ και 6.11 αναδεικνύουν ότι επιλέγονται λιγότερα δείγματα όσο εξελίσσεται η εκπαίδευση, που όμως περιέχουν μεγαλύτερο ποσοστό της σημαντικότητας όλων των δεδομένων. Παρατηρώντας ωστόσο όλα τα σενάρια, είναι ασφαλές να υποθέσουμε ότι στην παρούσα διάταξη μέσω της αξιολόγησης των δεδομένων η εκπαίδευση σταδιακά γίνεται σε σχεδόν όλα τα δεδομένα, και δίνεται έμφαση σε αυτά που εξακολοθούν να μην χειρίζονται με μεγαλύτερη ακρίβεια. Φαίνεται ωστόσο ότι με την επιλογή των σημαντικών δεδομένων ο υπολογιστικός φόρτος έχει τη δυνατότητα να προσαρμοστεί στις ανάγκες της εκπαίδευσης στο προσωπικό σύνολο δεδομένων, για να μην καταναλωθεί περισσότερη ενέργεια από την απαιτούμενη.

Στον πίνακα 6.1 παρουσιάζονται συγκεντρωτικά οι μέσοι όροι ενέργειας υπολογισμού και μετάδοσης για τις 10 εποχές, για κάθε στρατηγική ανάθεσης δεδομένων. Η ραγδαία αύξηση της ενέργειας υπολογισμού με την στρατηγική επιλογής ολόκληρου του συνόλου δεδομένων έγκειται στην λειτουργία περισσότερων επεξεργαστών σε υψηλότερες συχνότητες λειτουργίας για να είναι εφικτός ο χειρισμός τόσων δεδομένων μέσα στο χρονικό όριο. Ο αυξημένος χρόνος υπολογισμού οδηγεί επίσης στην ελάττωση του διαθέσιμου χρόνου για την μετάδοση, που οδηγεί στις συσκευές να μεταδίδουν σε μεγαλύτερη ισχύ, και άρα να καταναλώνουν περισσότερη ενέργεια. Αυτό αποτελεί ένα υποδειγματικό σενάριο για το πώς η ανάθεση δεδομένων, που αφορά ευθέως μόνο την ενέργεια υπολογισμού, μπορεί να επηρεάσει και την ενέργεια μετάδοσης, αφού αμφότερα βρίσκονται υπό τον κοινό χρονικό περιορισμό.

Ωστόσο, είναι εμφανές ότι η εκπαίδευση στο IID σύνολο δεδομένων οδηγεί σε πολύ γρήγορη σύγκλιση του μοντέλου, και ακόμα και αν ήδη αναδεικνύει επιθυμητά στοιχεία για τη λειτουργία του αλγορίθμου, για να καταστεί πληρέστερη η αξιολόγηση της στρατηγικής επιλογής δεδομένων θα πρέπει να αναδειχθεί περισσότερο η ποιότητα των δεδομένων που επιλέγονται. Γι'αυτό, θα εξετάσουμε και το σενάριο της εκπαίδευσης σε non-IID δεδομένα. Επειδή η εκπαίδευση είναι σαφώς δυσκολότερη, δίνονται παραπάνω γύροι επικοινωνίας για την σύγκλιση



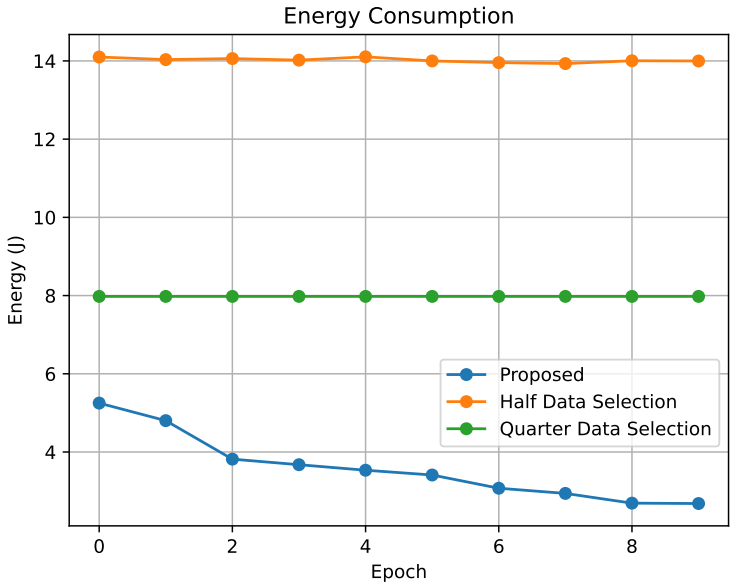
Σχήμα 6.14: Ακρίβεια που επιτυγχάνεται για κάθε στρατηγική επιλογής δεδομένων, σε iid δεδομένα



Σχήμα 6.15: Ακρίβεια που επιτυγχάνεται για κάθε στρατηγική επιλογής δεδομένων, σε non-iid δεδομένα

	Mean Energy Consumption per Epoch (J)		
Data Allocation Scheme	Computation	Communication	Total Energy
Proposed	2.301	1.257	3.558
Quarter	6.537	1.432	7.969
Half	12.485	1.541	14.026
Full	35.619	1.728	37.347

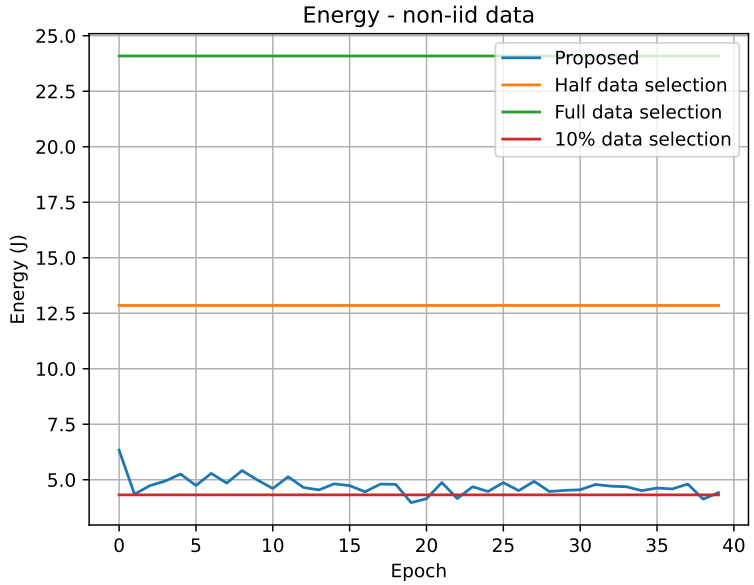
Πίνακας 6.1: Μέση ενέργεια υπολογισμού και επικοινωνίας για κάθε στρατηγική επιλογής δεδομένων



Σχήμα 6.16: Ενεργειακή κατανάλωση για κάθε στρατηγική επιλογής δεδομένων (iid-δεδομένα)

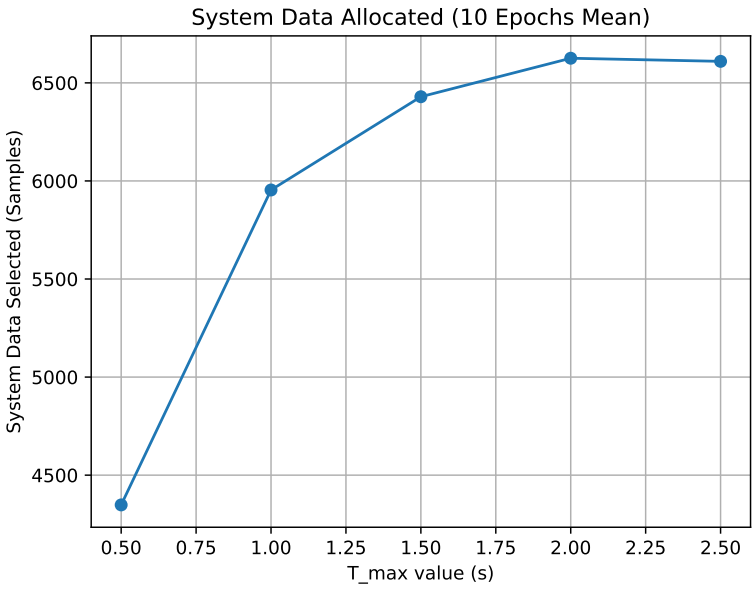
του μοντέλου. Επίσης, συμπεριλαμβάνεται το σενάριο της επιλογής του 10% των δεδομένων σε κάθε συσκευή, για να εξεταστεί η απόδοση ενός σεναρίου που επιλέγονται εν γένει λιγότερα δεδομένα από τον προτεινόμενο αλγόριθμο.

Στα σχήματα 6.15 και 6.17 φαίνεται η ακρίβεια που επιτυγχάνεται και η ενέργεια που καταναλώνεται για τις διάφορες στρατηγικές επιλογής δεδομένων στα non-IID δεδομένα. Παρατηρούμε αμέσως ότι η ενέργεια δεν μειώνεται τόσο αισθητά κατά την εκπαίδευση, κάτι που δεν αποτελεί έκπληξη εφόσον το διάγραμμα 6.12β΄ δείχνει ότι δεν ξεχωρίζουν ιδιαίτερα δεδομένα με μεγαλύτερη σημαντικότητα κατά την εκπαίδευση. Ωστόσο, πάλι φαίνεται να καταναλώνεται πολύ λιγότερη ενέργεια με παρόμοια απόδοση, όσον αφορά τις εναλλακτικές όπου επιλέγονται περισσότερα δεδομένα. Η ενέργεια που καταναλώνεται είναι ελαφρώς υψηλότερη από την επιλογή μόνο του 10% του συνόλου του δεδομένων, με αυτήν όμως να επιτυγχάνει μια ελαφρώς χαμηλότερη απόδοση στο τελικό μοντέλο. Γενικά οι αποδόσεις είναι εμφανώς χειρότερες σε σχέση με αυτές στα iid δεδομένα, κάτι που είναι πλήρως αναμενόμενο εφόσον αυτό αποτελεί γνωστό πρόβημα της Ομόσπονδης Μάθησης [30].

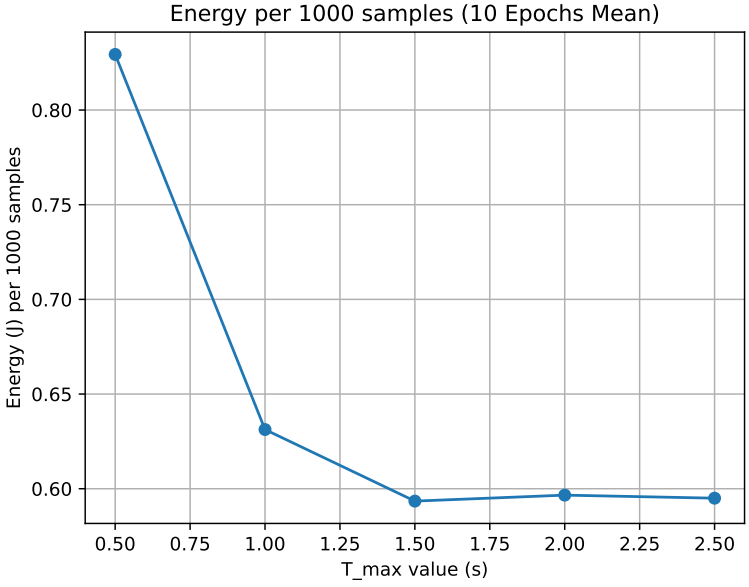


Σχήμα 6.17: Ενεργειακή κατανάλωση για κάθε στρατηγική επιλογής δεδομένων (non-iid δεδομένα)

Είναι επίσης αξιοσημείωτη η παρατήρηση ότι αυξάνοντας το όριο της διάρκειας ενός γύρου επικοινωνίας T_{max} αυξάνεται ο αριθμός των δεδομένων που αναθέτονται, ενώ μειώνεται η ενέργεια που καταναλώνεται ανά επιλεγμένο δείγμα, όπως φαίνεται από τις γραφικές παραστάσεις των σχημάτων 6.18 και 6.19. Η αύξηση του χρονικού ορίου T_{max} επιτρέπει τον χειρισμό περισσότερων δειγμάτων σε κάθε γύρο επικοινωνίας, που θεωρητικά οφείλει να οδηγήσει στην σύγκλιση του μοντέλου σε λιγότερους γύρους επικοινωνίας. Συνάμα, επιτρέπει την λειτουργία των επεξεργαστών σε χαμηλότερη συχνότητα, αν αυτό συνεισφέρει στη μείωση της ενέργειας, καθώς και στην μετάδοση των παραγώγων με λιγότερη ισχύ. Η διαμόρφωση στόχων και κριτηρίων για την βέλτιστη επιλογή του χρονικού ορίου T_{max} θα αποτελούσε μια πιθανή επέκταση της παρούσας εργασίας.



Σχήμα 6.18: Δεδομένα που επιλέγονται στο σύστημα για διαφορετικές τιμές του χρονικού ορίου T_{max} (Μέσος όρος 10 γύρων επικοινωνίας)



Σχήμα 6.19: Ενέργεια ανά 1000 δείγματα που επιλέγονται στο σύστημα (μέσος όρος 10 γύρων επικοινωνίας)

6.4 Αξιολόγηση της Στρατηγικής Κοινής Ανάθεσης Ραδιοπόρων και Υπολογιστικών Πόρων

Στο τελικό στάδιο της αξιολόγησης θα αντιπαραβάλουμε τον προτεινόμενο αλγόριθμο με τις εναλλακτικές στρατηγικές ανάθεσης πόρων που περιγράφηκαν στην υποενότητα 5.5.

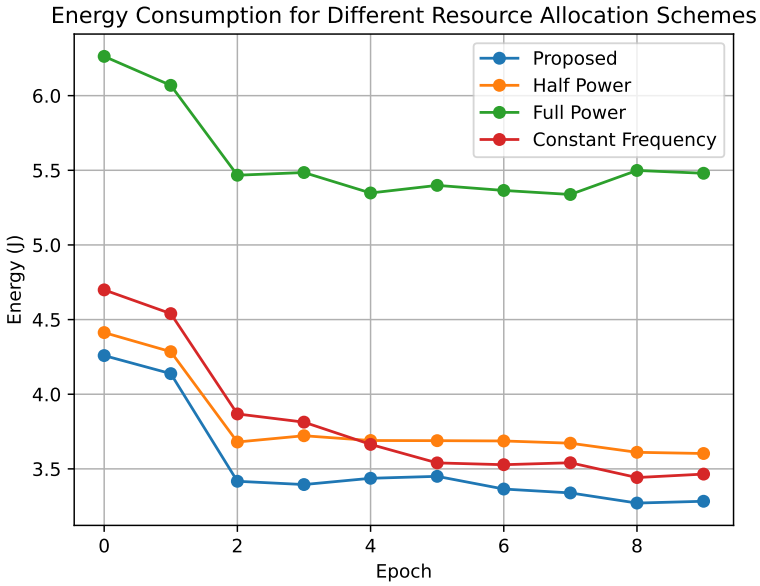
Θα παρακολουθήσουμε την ενεργειακή κατανάλωση του συστήματος στη διάρκεια ενός γύρου ομόπονδης μάθησης. Ο πίνακας 6.2, όπως και ο πίνακας 6.1 από την αξιολόγηση των στρατηγικών επιλογής δεδομένων παρουσιάζουν μια συσχέτιση ανάμεσα στην αύξηση της ενέργειας υπολογισμού και στην αντίστοιχη αύξηση της ενέργειας επικοινωνίας. Δηλαδή, όταν αντικαθιστούμε ένα συστατικό στοιχείο του αλγορίθμου που αφορά ευθέως την κατανάλωση υπολογισμού (ή της μετάδοσης αντίστοιχα) με μια πιο απλοϊκή στρατηγική, δεν είναι απαραίτητο ότι θα παρατηρηθούν καλύτερα αποτελέσματα για την ενέργεια μετάδοσης (ή αντίστροφα) για την οποία θα υπάρξει βελτιστοποίηση. Αυτό ενισχύει τη σημασία της ενιαίας βελτιστοποίησης των δεδομένων, συχνοτήτων, και ισχύων, επειδή όλα είναι μεταξύ τους συνδεδεμένα στο πρόβλημα που μας αφορά.

Η γραφική παράσταση 6.20 απεικονίζει την διατήρηση ελαφρώς χαμηλότερης ενέργειας υπολογισμού στη διάρκεια της ομόσπονδης μάθησης. Το γεγονός ότι δεν παρατηρείται σημαντική διαφορά μπορεί να σημαίνει ότι η βέλτιστη λύση δεν απέχει τόσο από τις υπόλοιπες στρατηγικές στα σχετικά τους συστατικά, όταν η στρατηγική επιλογής δεδομένων είναι κοινή, ωστόσο εξακολουθεί να αναδεικνύεται η βελτίωση της ενεργειακής απόδοσης όταν ο έλεγχος των ισχύων μετάδοσης και των συχνοτήτων επεξεργασίας βελτιστοποιούνται από κοινού, παρά ξεχωριστά.

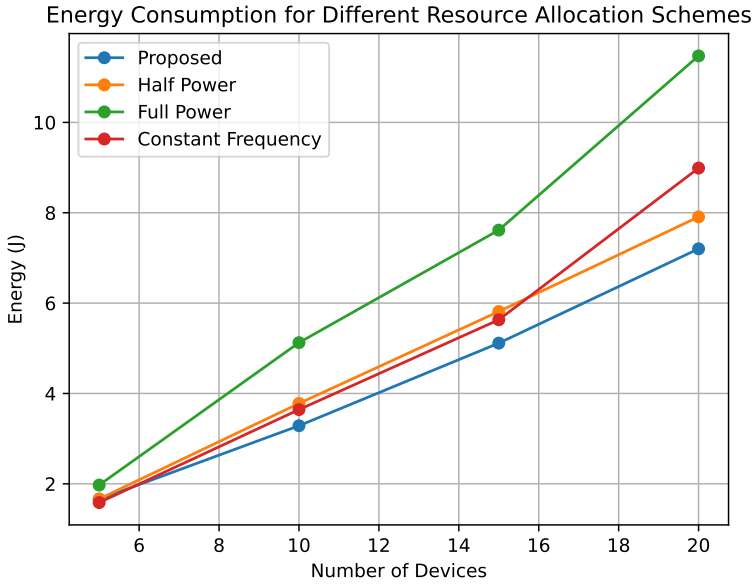
Το διάγραμμα 6.21 αναδεικνύει τα αντίστοιχα αποτελέσματα κατά την κλιμάκωση του συστήματος. Σημειώνεται ότι για αυτό το πείραμα το χρονικό όριο ενός γύρου επικοινωνίας T_{max} αυξήθηκε στο 1 δευτερόλεπτο, και σε κάθε συσκευή ανατέθηκαν 3.000 δείγματα. Η διαφορά στην ενέργεια που καταναλώνεται σε κάθε στρατηγική ανάθεσης πόρων γίνεται αρκετά πιο αισθητή όσο αυξάνεται ο αριθμός συσκευών, κάτι που αναδεικνύει την σημασία για την εφαρμογή ενεργειακά αποδοτικών τεχνικών ανάθεσης πόρων και δεδομένων σε μεγαλύτερα δίκτυα ομόσπονδης μάθησης.

	Mean Energy Consumption per Epoch (J)		
Resource Allocation Scheme	Computation	Communication	Total Energy
Proposed	2.321	1.231	3.552
Half Power	2.363	1.446	3.809
Full Power	3.970	1.597	5.567
Constant Frequencies	2.426	1.342	3.768

Πίνακας 6.2: Μέση ενέργεια υπολογισμού και επικοινωνίας για κάθε στρατηγική ανάθεσης πόρων.



Σχήμα 6.20: Ενέργεια κατανάλωσης κάθε στρατηγικής ανάθεσης πόρων κατά τη διάρκεια της ΟΜ



Σχήμα 6.21: Ενέργεια κατανάλωσης κατά την κλιμάκωση του δικτύου ΟΜ για διάφορες στρατηγικές ανάθεσης πόρων. Για κάθε διάταξη λαμβάνεται υπόψην ο μέσος όρος της ενέργειας στις 10 πρώτες εποχές.

7

Συμπεράσματα και Μελλοντικές Επεκτάσεις

Ο κεντρικός σκοπός αυτής της διπλωματικής εργασίας είναι η ενίσχυση της ενεργειακής αποδοτικότητας των ασύρματων δικτύων Ομόσπονδης Μάθησης, μέσω της από κοινού βελτιστοποίησης των ισχύων μετάδοσης των ανανεώσεων στις παραμέτρους, των συχνοτήτων λειτουργίας των επεξεργαστών, και των επιλεγμένων δεδομένων. Για αυτόν τον σκοπό χρησιμοποιούμε την ιδέα της σημαντικότητας των δεδομένων, προκειμένου να διατυπώσουμε μια στρατηγική επιλογής δεδομένων που θα επιταχύνουν την διαδικασία της μάθησης, και ανάθεσής τους στους επεξεργαστές προς τη μείωση της ενέργειας υπολογισμού.

Σύμφωνα με αυτήν την στρατηγική, μοντελοποιούμε μαθηματικά ένα πρόβλημα βελτιστοποίησης που εκφράζει τον στόχο της εκπαίδευσης όσων περισσότερων σημαντικών δεδομένων με τη λιγότερη δυνατή ενέργεια υπό ορισμένες συνθήκες. Για την επίλυσή του επιστρατεύουμε τεχνικές Κλασικής Βελτιστοποίησης που καθιστούν το πρόβλημα αυτό διαχειρίσιμο και εξασφαλίζουν την αποδοτική εύρεση μιας υποβέλτιστης λύσης. Αναπτύσσουμε έτσι έναν αλγόριθμο βελτιστοποίησης πόρων και δεδομένων με βάση τη σημαντικότητα, που αποτελεί την βασική συνεισφορά αυτής της εργασίας. Μέσω της εκτενούς πειραματικής μελέτης αναδεικνύεται η ενεργειακή αποδοτικότητα του προτεινόμενου αλγορίθμου, μέσω της αντιπαραβολής του με εναλλακτικές πιθανές στρατηγικής ανάθεσης πόρων και δεδομένων.

Η παρούσα εργασία επιδέχεται τις ακόλουθες επεκτάσεις

- Το χρονικό όριο ενός γύρου επικοινωνίας θεωρείται σταθερό ως ένας QoS περιορισμός. Ωστόσο, θα μπορούσε να τεθεί προς βελτιστοποίηση προκειμένου εκτός από την ενέργεια κατανάλωσης να ελαττωθεί και ο χρόνος του γύρου επικοινωνίας, μέσω της ενσωμάτωσής του σε αντικειμενική συνάρτηση ενός νέου προβλήματος βελτιστοποίησης.

- Η μετρική της σημαντικότητας των δεδομένων είναι ιδιαίτερα ευάλωτη σε κίβδηλα δεδομένα, όπως δεδομένα εκπαίδευσης με λανθασμένες ταμπέλες. Προς αυτήν την καυεύθυνση θα συνέβαλε η χρήση της Τιμής Shapley προς την αναγνώριση συμμετεχόντων που ενδεχομένως προσπαθούν να υπονομεύσουν τη διαδικασία της ομόσπονδης μάθησης με λανθασμένα δεδομένα, ενώ θα χρησίμευε και γενικότερα για την εκτίμηση της συνεισφοράς κάθε χρήστη στην βελτίωση της απόδοσης του μοντέλου [42].

- Τέλος, στην παρούσα διπλωματική δεν λαμβάνεται κάποια μέριμνα για την αντιμετώπιση των προβλημάτων που επιφέρουν τα non-iid δεδομένα στην ομόσπονδη μάθηση. Μια πιθανή λύση σε αυτό θα ήταν η επέκταση του προτεινόμενου αλγορίθμου σε ασύρματα δίκτυα ιεραρχικής ομόσπονδης μάθησης (Hierarchical Federated Learning - HFL), που εκτός από τον καλύτερο χειρισμό των non-iid δεδομένων αποτρέπει την κεντρική οντότητα από το να αποτελεί στενωπό (bottleneck) για την διαδικασία της ομόσπονδης μάθησης.

Βιβλιογραφία

[1] E. B. Bajalinov. *Linear-Fractional Programming: Theory, Methods, Applications and Software*. Kluwer Academic Publishers, 2003.

[2] Eli Bendersky. *The Softmax Function and Its Derivative*. 2016. URL: https://eli.thegreenplace.net/2016/the-softmax-function-and-its-derivative/.

[3] James C. Bezdek, Chris Coray, Robert Gunderson και James Watson. «Detection and Characterization of Cluster Substructure II. Fuzzy c-Varieties and Convex Combinations Thereof». Στο: *SIAM Journal on Applied Mathematics* 40.2 (1981), σσ. 358–372. DOI: 10.1137/0140030. eprint: https://doi.org/10.1137/0140030. URL: https://doi.org/10.1137/0140030.

[4] James C. Bezdek και Richard J. Hathaway. «Some Notes on Alternating Optimization». Στο: AFSS '02. Berlin, Heidelberg: Springer-Verlag, 2002, σσ. 288–300. ISBN: 3540431500.

[5] Stephen Boyd και Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[6] Tianyi Chen, Georgios B. Giannakis, Tao Sun και Wotao Yin. *LAG: Lazily Aggregated Gradient for Communication-Efficient Distributed Learning*. 2018. arXiv: 1805.09965 [stat.ML]. URL: https://arxiv.org/abs/1805.09965.

[7] A. R. Conn, N. I. Gould και P. L. Toint. *Trust Region Methods*. SIAM, 2000, σ. 19.

[8] TensorFlow Datasets. *MNIST dataset*. https://www.tensorflow.org/datasets/catalog/mnist. Accessed: 2024-09-19. 2024.

[9] Maria Diamanti, Christos Pelekis, Eirini Eleni Tsiropoulou και Symeon Papavassiliou. «Delay Minimization for Rate-Splitting Multiple Access-Based Multi-Server MEC Offloading». Στο: *IEEE/ACM Transactions on Networking* 32.2 (2024), σσ. 1035–1047. DOI: 10.1109/TNET.2023.3311131.

[10] W. Dinkelbach. «On nonlinear fractional programming». Στο: *Management Science* 13.7 (Μαρ. 1967), σσ. 492–498.

[11] Ian Goodfellow, Yoshua Bengio και Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[12] Michael T. Goodrich και Roberto Tamassia. *Algorithm Design: Foundations, Analysis, and Internet Examples*. John Wiley & Sons, 2002. Κεφ. 5.1.1 The Fractional Knapsack Problem, σσ. 259–260.

[13] Andrew Hard, Kanishka Rao, Rajiv Mathews, Swaroop Ramaswamy, Françoise Beaufays, Sean Augenstein, Hubert Eichner, Chloé Kiddon και Daniel Ramage. *Federated Learning for Mobile Keyboard Prediction*. 2019. arXiv: 1811.03604 [cs.CL]. URL: https://arxiv.org/abs/1811.03604.

[14] Chaofan He, Yang Hu, Yan Chen και Bing Zeng. «Joint Power Allocation and Channel Assignment for NOMA With Deep Reinforcement Learning». Στο: *IEEE Journal on Selected Areas in Communications* 37.10 (2019), σσ. 2200–2210. DOI: 10.1109/JSAC.2019.2933762.

[15] Yinghui He, Jinke Ren, Guanding Yu και Jiantao Yuan. «Importance-Aware Data Selection and Resource Allocation in Federated Edge Learning System». Στο: *IEEE Transactions on Vehicular Technology* 69.11 (2020), σσ. 13593–13605. DOI: 10.1109/TVT.2020.3015268.

[16] Kenichi Higuchi και Anass Benjebbour. «Non-orthogonal Multiple Access (NOMA) with Successive Interference Cancellation for Future Radio Access». Στο: *IEICE Transactions on Communications* E98.B (Μαρ. 2015), σσ. 403–414. DOI: 10.1587/transcom.E98.B.403.

[17] Marius Hobbhahn και Jaime Sevilla. *What's the Backward-Forward FLOP Ratio for Neural Networks?* Accessed: 2024-09-02. 2021. URL: https://epochai.org/blog/backward-forward-FLOP-ratio.

[18] S. Islam, N. Avazov, O. Dobre και K. Kwak. «Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges». Στο: *IEEE Communications Surveys Tutorials* 19 (2016), σσ. 721–742. DOI: 10.1109/COMST.2016.2621116.

[19] Angelos Katharopoulos και François Fleuret. *Not All Samples Are Created Equal: Deep Learning with Importance Sampling*. 2019. arXiv: 1803.00942 [cs.LG]. URL: https://arxiv.org/abs/1803.00942.

[20] Mykel J. Kochenderfer και Tim A. Wheeler. *Algorithms for Optimization*. Cambridge, MA: MIT Press, 2019. ISBN: 9780262039420.

[21] Thomas Kurbiel. *Derivative of the Softmax Function and the Categorical Cross-Entropy Loss*. 2019. URL: https://towardsdatascience.com/derivative-of-the-softmax-function-and-the-categorical-cross-entropy-loss-ffceefc081d1.

[22] Yann LeCun, Yoshua Bengio και Geoffrey Hinton. «Deep learning». Στο: *Nature* 521 (2015), σσ. 436–444.

[23] Yann LeCun, Corinna Cortes και CJ Burges. «MNIST handwritten digit database». Στο: *ATT Labs [Online]* (2010). Accessed: 2024-09-19.

[24] Youngmoon Lee, Kang G. Shin και Hoon Sung Chwa. «Thermal-Aware Scheduling for Integrated CPUs--GPU Platforms». Στο: *ACM Trans. Embed. Comput. Syst.* 18.5s (Οκτ. 2019). ISSN: 1539-9087. DOI: 10.1145/3358235. URL: https://doi.org/10.1145/3358235.

[25] Qiuwei Li, Zhihui Zhu και Gongguo Tang. «Alternating Minimizations Converge to Second-Order Optimal Solutions». Στο: *Proceedings of the 36th International Conference on Machine Learning*. Επιμέλεια υπό Kamalika Chaudhuri και Ruslan Salakhutdinov. Τόμ. 97. Proceedings of Machine Learning Research. PMLR, Σεπτ. 2019, σσ. 3935–3943. URL: https://proceedings.mlr.press/v97/li19n.html.

[26] Cong Liu, Jian Li, Wei Huang, Juan Rubio, Evan Speight και Felix Xiaozhu Lin. «Power-efficient time-sensitive mapping in heterogeneous systems». Στο: *2012 21st International Conference on Parallel Architectures and Compilation Techniques (PACT)*. 2012, σσ. 23–32.

[27] Dongzhu Liu, Guangxu Zhu, Jun Zhang και Kaibin Huang. *Data-Importance Aware User Scheduling for Communication-Efficient Edge Machine Learning*. 2019. arXiv: 1910.02214 [cs.NI]. URL: https://arxiv.org/abs/1910.02214.

[28] Yuanwei Liu, Zhijin Qin, Maged Elkashlan, Zhiguo Ding, Arumugam Nallanathan και Lajos Hanzo. «Nonorthogonal Multiple Access for 5G and Beyond». Στο: *Proceedings of the IEEE* 105.12 (2017), σσ. 2347–2381. DOI: 10.1109/JPROC.2017.2768666.

[29] Kun Lu, Zhanji Wu και Xuanbo Shao. «A Survey of Non-Orthogonal Multiple Access for 5G». Στο: *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)* (2017), σσ. 1–5. DOI: 10.1109/VTCFall.2017.8288400.

[30] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson και Blaise Agüera y Arcas. *Communication-Efficient Learning of Deep Networks from Decentralized Data*. 2023. arXiv: 1602.05629 [cs.LG]. URL: https://arxiv.org/abs/1602.05629.

[31] John von Neumann. «Über ein ökonomisches Gleichgewichtssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes». Στο: *Ergebnisse eines Mathematischen Kolloquiums* 8 (1937), σσ. 73–83.

[32] Dinh C. Nguyen, Viet Quoc Pham, P. Pathirana, Ming Ding, A. Seneviratne, Zihuai Lin, O. Dobre και W. Hwang. «Federated Learning for Smart Healthcare: A Survey». Στο: *ACM Computing Surveys (CSUR)* 55 (2021), σσ. 1–37. DOI: 10.1145/3501296.

[33] Xianke Qiang, Yun Hu, Zheng Chang και Timo Hamalainen. «Importance-aware data selection and resource allocation for hierarchical federated edge learning». Στο: *Future Generation Computer Systems* 154 (2024), σσ. 35–44. ISSN: 0167-739X. DOI: https://doi.org/10.1016/j.future.2023.12.014. URL: https://www.sciencedirect.com/science/article/pii/S0167739X23004740.

[34] Lutz Roeder. *Netron: Visualizer for neural network, deep learning and machine learning models*. https://github.com/lutzroeder/netron. Accessed: 2024-09-18. 2024.

[35] Lutz Roeder. *Netron: Visualizer for neural network, deep learning and machine learning models*. https://netron.app/. Accessed: 2024-09-18. 2024.

[36] Rukhsana Ruby, Hailiang Yang, Felipe A. P. de Figueiredo, Thien Huynh-The και Kaishun Wu. «Energy-Efficient Multiprocessor-Based Computation and Communication Resource Allocation in Two-Tier Federated Learning Networks». Στο: *IEEE Internet of Things Journal* 10.7 (2023), σσ. 5689–5703. DOI: 10.1109/JIOT.2022.3153996.

[37] S. Schaible. «Fractional programming». Στο: *Zeitschrift für Operations Research* 27 (Οκτ. 1982), σσ. 39–54.

[38] SciPy. *Trust-Region Constrained Algorithm*. Accessed: 2024-09-19. 2024. URL: https://docs.scipy.org/doc/scipy/reference/optimize.minimize-trustconstr.html.

[39] Kaiming Shen και Wei Yu. «Fractional Programming for Communication Systems—Part I: Power Control and Beamforming». Στο: *IEEE Transactions on Signal Processing* 66.10 (2018), σσ. 2616–2630. DOI: 10.1109/TSP.2018.2812733.

[40] Lingyang Song, Yonghui Li, Zhiguo Ding και H. Vincent Poor. *Resource Management in Non-orthogonal Multiple Access Networks for 5G and Beyond*. 2017. arXiv: 1610.09465 [cs.IT]. URL: https://arxiv.org/abs/1610.09465.

[41] I. M. Stancu-Minasian. *Fractional Programming: Theory, Methods and Applications*. Kluwer Academic Publishers, 1992.

[42] Qiheng Sun, Xiang Li, Jiayao Zhang, Li Xiong, Weiran Liu, Jinfei Liu, Zhan Qin και Kui Ren. «ShapleyFL: Robust Federated Learning Based on Shapley Value». Στο: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. KDD '23. Long Beach, CA, USA: Association for Computing Machinery, 2023, σσ. 2096–2108. ISBN: 9798400701030. DOI: 10.1145/3580305.3599500. URL: https://doi.org/10.1145/3580305.3599500.

[43] Z. Wang. *Traditional Federated Learning*. https://github.com/wzljerry/Hierarchical-Federated-Learning/blob/main/FL.ipynb. Accessed: 2024-09-18. 2022.

[44] John H. Wolfe. «PATTERN CLUSTERING BY MULTIVARIATE MIXTURE ANALYSIS». Στο: *Multivariate Behavioral Research* 5.3 (1970). PMID: 26812701, σσ. 329–350. DOI: 10.1207/s15327906mbr0503_6. eprint: https://doi.org/10.1207/s15327906mbr0503_. URL: https://doi.org/10.1207/s15327906mbr0503_6.

[45] Zhanji Wu, Kun Lu, Chengxin Jiang και Xuanbo Shao. «Comprehensive Study and Comparison on 5G NOMA Schemes». Στο: *IEEE Access* 6 (2018), σσ. 18511–18519. DOI: 10.1109/ACCESS.2018.2817221.

[46] Zhaohui Yang, Mingzhe Chen, Walid Saad, Choong Seon Hong και Mohammad Shikh-Bahaei. *Energy Efficient Federated Learning Over Wireless Communication Networks*. 2020. arXiv: 1911.02417 [cs.IT]. URL: https://arxiv.org/abs/1911.02417.

[47] Qunsong Zeng, Yuqing Du, Kaibin Huang και Kin K. Leung. «Energy-Efficient Radio Resource Allocation for Federated Edge Learning». Στο: *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*. 2020, σσ. 1–6. DOI: 10.1109/ICCWorkshops49005.2020.9145118.

[48] Jun Zhang, Lipeng Zhu, Zhenyu Xiao, Xianbin Cao, Dapeng Oliver Wu και Xiang-Gen Xia. «Optimal and Sub-Optimal Uplink NOMA: Joint User Grouping, Decoding Order, and Power Control». Στο: *IEEE Wireless Communications Letters* 9.2 (2020), σσ. 254–257. DOI: 10.1109/LWC.2019.2951765.

[49] Xinran Zhang, Zhimin He, Yaohua Sun, Shuo Yuan και Mugen Peng. «Joint Sensing, Communication, and Computation Resource Allocation for Cooperative Perception in Fog-Based Vehicular Networks». Στο: *2021 13th International Conference on Wireless Communications and Signal Processing (WCSP)*. 2021, σσ. 1–6. DOI: 10.1109/WCSP52459.2021.9613157.

[50] Guangyuan Zheng, Chen Xu, Hao Long και Xiongwen Zhao. «MEC in NOMA-HetNets: A Joint Task Offloading and Resource Allocation Approach». Στο: *2021 IEEE Wireless Communications and Networking Conference (WCNC)*. 2021, σσ. 1–6. DOI: 10.1109/WCNC49053.2021.9417280.

[51] Guangxu Zhu, Dongzhu Liu, Yuqing Du, Changsheng You, Jun Zhang και Kaibin Huang. «Toward an Intelligent Edge: Wireless Communication Meets Machine Learning». Στο: *IEEE Communications Magazine* 58.1 (2020), σσ. 19–25. DOI: 10.1109/MCOM.001.1900103.

Γλωσσάριο

Απόδοση	Ξενόγλωσσος όρος
ανάθεση ισχύος	power allocation
Πληροφορίες Κατάστασης Καναλιού	Channel State Information (CSI)
Ενεργειακή Απόδοση	Energy Efficiency
Διαδίκτυο των Πραγμάτων	Internet of Things (IoT)
Μη Ορθογωνική Πολλαπλή Πρόσβαση	Non Orthogonal Multiple Access (NOMA)
Σταθμός Βάσης	Base Station (BS)
κατώφλι	threshold
ανερχόμενη ζεύξη	uplink
κατερχόμενη ζεύξη	downlink
Διαδοχική Ακύρωση Παρεμβολών	Successive Interference Cancellation (SIC)
Λευκός Προσθετικός Θόρυβος Γκάους	Additive White Gaussian Noise (AWGN)
ανοδική/καθοδική κλίση	gradient ascent/descent
Σταθμισμένο Μέσο Τετραγωνικό Σφάλμα	Weighted Mean Square Error (WMSE)
Συνελικτικό Νευρωνικό Δίκτυο	CNN
Σημαντικότητα Δεδομένων	Data Importance
Ομόσπονδη Μάθηση	Federated Learning
διακομιστής/εξυπηρετητής	server
Κλασματικός Προγραμματισμός	Fractional Programming
Κυρτή Βελτιστοποίηση	Convex Optimization
Εναλασσόμενη Βελτιστοποίηση	Alternating Optimization
Ποιότητα Υπηρεσίας	Quality of Service
Πολλαπλασιαστές Lagrange	Lagrange Multipliers
Εμπρόσθια Διάδοση	Forward Pass
Οπισθοδιάδοση	Backpropagation
συνάρτηση απώλειας	loss function
συνάρτηση ενεργοποίησης	activation function
επίπεδο εξόδου	output layer

ταμπέλα	label
Ιακωβιανός πίνακας	Jacobian matrix
ανά σημείο ελάχιστο	pointwise minimum
υποπαράγωγος	subgradient
πράξεις κινητής υποδιαστολής	floating point operations
γύρος επικοινωνίας	communication round
φασματική πυκνότητα ισχύος	power spectral density