



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ
ΕΡΓΑΣΤΗΡΙΟ ΟΡΑΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ, ΕΠΙΚΟΙΝΩΝΙΑΣ ΛΟΓΟΥ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ
ΖΩΓΡΑΦΟΥ 157 73, ΑΘΗΝΑ

Τροπική Ανάλυση Αλγορίθμων σε Γράφους

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

Εμμανουήλ Θεοδώση

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

ΕΡΓΑΣΤΗΡΙΟ ΟΡΑΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ, ΕΠΙΚΟΙΝΩΝΙΑΣ ΛΟΓΟΥ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ
Αθήνα, Ιούλιος 2018



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής
Εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας
Σημάτων
Ζωγράφου 157 73, Αθήνα

Τροπική Ανάλυση Αλγορίθμων σε Γράφους

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

Εμμανουήλ Θεοδώση

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 20 Ιουλίου, 2018.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Πέτρος Μαραγκός
Καθηγητής
Ε.Μ.Π.

.....
Γεράσιμος Ποταμιάνος
Αναπληρωτής Καθηγητής
Πανεπιστήμιο Θεσσαλίας

.....
Κωνσταντίνος Τζαφέστας
Αναπληρωτής Καθηγητής
Ε.Μ.Π.

Αθήνα, Ιούλιος 2018

(Υπογραφή)

.....
Εμμανουήλ Θεοδώσης

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

Copyright ©–All rights reserved Εμμανουήλ Θεοδώσης, 2018.

Με επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν το συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Στη γιαγιά μου, Ειρήνη, που δεν πρόλαβε να με δει πτυχιούχο.

Ευχαριστίες

Ολοκληρώνοντας αυτόν τον μεγάλο κύκλο στην πορεία της ζωής μου, φτάνω σε ένα από τα πιο σημαντικά στάδια για εμένα: το στάδιο όπου ευχαριστώ, ειλικρινά και εγκάρδια, τους ανθρώπους που διαμόρφωσαν το ποιόν μου, που μου στάθηκαν όλα αυτά τα χρόνια, και που με στήριξαν παραπάνω από όσο μπορούσαν.

Καταρχάς, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή της παρούσας εργασίας, κ. Πέτρο Μαραγκό. Θα ήθελα να τον ευχαριστήσω για την εμπιστοσύνη του, τον χρόνο που μου διέθεσε, το ενδιαφέρον του, τις συμβουλές του, και την καθοδήγηση του. Αλλά πάνω απ' όλα, θα ήθελα να τον ευχαριστήσω που πάντα με παρότρυνε να αναρωτηθώ. Από τα πρώτα φοιτητικά μου χρόνια, ενθάρρυνε τον ενθουσιασμό μου για την έρευνα και τα μαθηματικά, προτρέποντας με να ψάξω, να παρατηρήσω, και να δοκιμάσω.

Έπειτα, θα ήθελα να ευχαριστήσω την οικογένειά μου. Τους ανθρώπους που υποστήριξαν τα όνειρά μου, που με βοηθούσαν στα εύκολα και στα δύσκολα, πάντα χωρίς προϋποθέσεις και απαιτήσεις. Θέλω να τους ευχαριστήσω για όσα μου προσέφεραν, θυσιάζοντας το υστέρημά τους ώστε να επιτρέψουν σε εμένα να κυνηγήσω τους στόχους και τις επιθυμίες μου.

Θα ήθελα να ευχαριστήσω τους φίλους μου, που ήταν εκεί, που με παρότρυναν, και με βοηθούσαν να ηρεμώ και να θυμάμαι ότι υπάρχουν και άλλα πράγματα στη ζωή πέραν του διαβάσματος. Θα ήθελα να ευχαριστήσω ιδιαίτερα την Ξένια, η οποία ανέχτηκε τις ιδιοτροπίες μου όλα αυτά τα χρόνια, και πάντα άκουγε τα παράπονά μου, προσφέροντας συμβουλές, και δίνοντας μου την ειλικρινή της γνώμη, όσο ωμή και αν ήταν, όταν τη χρειαζόμουν περισσότερο.

Τέλος, θα ήθελα να ευχαριστήσω όλους τους ανθρώπους οι οποίοι πέρασαν από τη ζωή μου. Είτε μου προσέφεραν θετικές είτε αρνητικές αναμνήσεις, τους είμαι ευγνώμων. Γιατί αυτές αλληλεπιδράσεις και οι εμπειρίες διαμόρφωσαν τον άνθρωπο τον οποίο είμαι, τις πεποιθήσεις μου, τις σκέψεις μου, τη συμπεριφορά μου, και δεν θα μπορούσα να είμαι για κάτι περισσότερο ευγνώμων από τον άνθρωπο ο οποίος είμαι σήμερα. Οπότε σας ευχαριστώ, όπου και αν είστε γιατί βοηθήσατε να γίνω αυτό που είμαι.

Εμμανουήλ Θεοδώσης,
Ιούλιος 2018

Περίληψη

Η μηχανική μάθηση αποτελεί ένα πεδίο το οποίο έχει βιώσει ραγδαία ανάπτυξη τα τελευταία χρόνια, με το ενδιαφέρον των ερευνητών να κορυφώνεται. Όμως, οι μέθοδοι που αναπτύσσονται συχνά βασίζονται σε ευριστικές οι οποίες δεν έχουν ξεκάθαρα θεμελιωμένο μαθηματικό υπόβαθρο, και συνεπώς υστερούν τόσο από πλευράς διαίσθησης, εννοώντας ότι δεν είναι διαφανείς οι λόγοι που έχουν τις εκάστοτε αποδόσεις, αλλά υστερούν και από πλευράς επεκτασιμότητας, καθώς δεν είναι εμφανές τι ιδέες θα μπορούσαν να εφαρμοστούν για την βελτίωση των μοντέλων.

Τα τελευταία χρόνια έχει υπάρξει μια τάση εισαγωγής αυστηρών μαθηματικών φορμαλισμών και γεωμετρικών στοιχείων στη μηχανική μάθηση. Σε αυτό το πνεύμα, αυτή η εργασία καταπιάνεται με την μοντελοποίηση αλγορίθμων που δρουν σε Weighted Finite State Transducers, τα οποία χρησιμοποιούνται συχνά στην Αυτόματη Αναγνώριση Φωνής και την Επεξεργασία Φυσικής Γλώσσας, στο πεδίο της τροπικής άλγεβρας. Μέσω αυτής της μοντελοποίησης αποσκοπούμε στην προσεκτικότερη μελέτη των ιδιοτήτων τους, και, όπου είναι δυνατό, επιθυμούμε να εκμεταλλευτούμε στοιχεία της τροπικής γεωμετρίας για να αναλύσουμε τους εν λόγω αλγορίθμους.

Συγκεκριμένα, στα πλαίσια της διπλωματικής εργασίας, ασχοληθήκαμε με την μοντελοποίηση των παρακάτω αλγορίθμων:

- αλγόριθμος προώθησης βαρών,
- αλγόριθμος αφαίρεσης ε -μεταβάσεων,
- αλγόριθμος Viterbi,
- και αλγόριθμος Viterbi με αποκοπή.

Μετά το πέρας της μοντελοποίησης, εκμεταλλευόμαστε στοιχεία της τροπικής γεωμετρίας, όπου αυτά εφαρμόζονται, έτσι ώστε να ορίσουμε γεωμετρικές οντότητες στον χώρο λύσεων των αλγορίθμων, προσφέροντας μια καινούργια οπτική, και εισάγουμε μετρικές οι οποίες μπορούν να χρησιμοποιηθούν στο σχεδιασμό συστημάτων.

Τέλος, εφαρμόζουμε τις προτεινόμενες μεθόδους σε σύνολα δεδομένων και αξιολογούμε τα αποτελέσματα, παρουσιάζοντας, παράλληλα, παραδείγματα που υπογραμμίζουν την κομψότητα της μοντελοποίησης στον κύριο κορμό του κειμένου.

Λέξεις Κλειδιά

Τροπική Άλγεβρα, Τροπική Γεωμετρία, Μαθηματική Μοντελοποίηση, Μηχανική Μάθηση, Βελτιστοποίηση

Abstract

Machine learning is a field that has experienced explosive growth in recent years, with researchers' interest on the field peaking. However, the methods developed are frequently based on heuristics which lack an established mathematical background, and thus suffer from the perspective of perception, meaning that the reasons for their performance are not transparent, but they also suffer from an extensibility standpoint, since it is not clear what ideas could be employed to further improve the models.

In the recent years there has been a trend of injecting mathematical modeling and geometrical elements into machine learning. In that spirit, this thesis devotes itself to the modeling of algorithms who act on Weighted Finite State Transducers, which are frequently used in Automatic Speech Recognition and Natural Language Processing, in the field of tropical algebra. Through this modeling we aim to meticulously study their properties, and, wherever possible, we wish to employ elements of tropical geometry to analyse these algorithms.

Specifically, during this diploma thesis, we devoted ourselves to the modeling of the following algorithms:

- the weight pushing algorithm,
- the epsilon removal algorithm,
- the Viterbi algorithm,
- and the pruning variant of the Viterbi algorithm.

After we have concluded modeling the algorithms, we utilize elements of tropical geometry, wherever they are applicable, so that we can define geometrical entities on the solution space of the algorithms, offering a new perspective, and we introduce certain metrics which can be used in the designing stage of such systems.

Finally, we apply the proposed methods to data sets and we evaluate the results, while also presenting numerical examples which underline the elegance of the modeling in the main body of the text.

Keywords

Tropical Algebra, Tropical Geometry, Mathematical Modeling, Machine Learning, Optimisation

Περιεχόμενα

Ευχαριστίες	11
Περίληψη	15
Abstract	17
Περιεχόμενα	20
Κατάλογος Σχημάτων	21
Κατάλογος Πινάκων	23
1 Εισαγωγή	27
1.1 Μηχανική Μάθηση	27
1.1.1 Βασικές ιδέες	27
1.1.2 Κριτήρια ελαχιστοποίησης	29
1.1.3 Διεργασίες μηχανικής μάθησης	31
1.2 Προβλήματα μηχανικής μάθησης	33
1.3 Στόχος της διπλωματικής εργασίας	34
1.4 Συμβολισμός	35
2 Τροπική Άλγεβρα και Γεωμετρία	39
2.1 Γραμμική Άλγεβρα	39
2.2 Τροπική Άλγεβρα	39
2.3 Τροπική Γεωμετρία	44
3 Γραφικά Μοντέλα	51
3.1 Γράφοι	51
3.2 Hidden Markov Models	54
3.3 Weighted Finite State Transducers	57
4 Τροπικοί Αλγόριθμοι σε WFST	61
4.1 Αλγόριθμος προώθησης βαρών	62
4.1.1 Ανάλυση	63
4.2 Αλγόριθμος αφαίρεσης ε -μεταβάσεων	69
4.2.1 Ανάλυση	69
4.3 Αλγόριθμος Viterbi	78
4.3.1 Ανάλυση	79
4.4 Αλγόριθμος Viterbi με αποκοπή	83
4.4.1 Ανάλυση	84

4.4.2	Γεωμετρία	85
4.4.3	Ερμηνεία	87
5	Πειραματισμός	93
5.1	Πειράματα σε Επεξεργασία Φυσικής Γλώσσας	93
5.1.1	Συμπεράσματα	96
5.2	Πειράματα με προσομοίωση σε IP Spoofing	97
5.2.1	Θεωρητικό Υπόβαθρο	99
5.2.2	Πείραμα 1: Εντοπισμός αποκλίνουσας κατάστασης	101
5.2.3	Πείραμα 2: Εντοπισμός εναλλάσσουσας αποκλίνουσας κατάστασης	105
5.2.4	Πείραμα 3: Δυναμική εναλλαγή παραμέτρου χαλάρωσης	109
6	Επίλογος	121
6.1	Συνεισφορά	121
6.2	Μελλοντική Έρευνα	121

Κατάλογος Σχημάτων

1.1	Γραμμική υπόθεση για την αλληλεπίδραση μεταξύ δύο ποσοτήτων και προσαρμογή της καμπύλης στα δεδομένα.	28
2.1	Παράδειγμα εύρεσης ελάχιστου μονοπατιού.	41
2.2	Αναπαράσταση ενός Ευκλείδειου ημιχώρου, ενός πολυέδρου, και ενός πολυτόπου.	44
2.3	Αναπαράσταση της γενικής εξίσωσης της ευθείας στην αλγεβρική και τροπική γεωμετρία.	45
2.4	Αναπαράσταση της εξίσωσης παραβολής στην αλγεβρική και τροπική γεωμετρία.	47
3.1	Ένας τυπικός γράφος.	52
3.2	Γενική μορφή ενός HMM.	54
3.3	HMM για την έκφραση ασθένειας.	56
3.4	WFST για μετατροπή πεζών σε κεφαλαία.	58
4.1	Παράδειγμα χρησιμότητας του αλγορίθμου προώθησης βαρών.	62
4.2	Δίκτυο για το παράδειγμα του αλγορίθμου προώθησης βαρών.	67
4.3	Βοηθητική επεξήγηση για τον Αλγόριθμο 2.	70
4.4	Βοηθητική επεξήγηση της χρησιμότητας του πίνακα E	73
4.5	Δίκτυο για το παράδειγμα του αλγορίθμου αφαίρεσης ϵ -μεταβάσεων.	75
4.6	Προκύπτων δίκτυο μετά την εφαρμογή του αλγορίθμου αφαίρεσης ϵ -μεταβάσεων.	76
4.7	Προκύπτων δίκτυο μετά την εφαρμογή του αλγορίθμου ψαλιδίσματος.	77
4.8	Συμβολικό παράδειγμα εφαρμογής του αλγορίθμου Viterbi.	79
4.9	Δίκτυο για το παράδειγμα εφαρμογής του αλγορίθμου Viterbi.	80
4.10	Δυνατές αναθέσεις περιορισμένες από τον νόμο ανανέωσης του Viterbi (κατώτεροι και αριστερότεροι περιορισμοί) και από την διαδικασία αποκοπής (άνωτεροι και δεξιότεροι περιορισμοί).	87
4.11	Τα προκύπτων πολύτοπα του Παραδείγματος 4.4.	90
5.1	Transliteration από ένα αλφάβητο α (λατινικό) σε ένα αλφάβητο β (ελληνικό).	94
5.2	Αιτήματα χρηστών σε έναν κεντρικό υπολογιστή. Με κόκκινο έχει επισημειωθεί ο κακόβουλος χρήστης που εκτελεί αιτήματα με σημαντικά μεγαλύτερο ρυθμό.	101
5.3	Εντοπισμός αποκλίνουσας κατάστασης: Ο κακόβουλος χρήστης παραμένει σταθερός μεταξύ των χρονικών παραθύρων.	101
5.4	Εντοπισμός εναλλάσσουσας αποκλίνουσας κατάστασης: Ο κακόβουλος χρήστης δύναται να αλλάζει μεταξύ των χρονικών παραθύρων.	106
5.5	Αυτόματη προσαρμογή παραμέτρου χαλάρωσης θ : Η επισημειωμένη κατάσταση σε κάθε χρονικό πλαίσιο υποδηλώνει την κατάσταση μεγίστης πιθανότητας.	110

Κατάλογος Πινάκων

2.1	Χαρακτηριστικές διαφορές θεμελιωδών στοιχείων μεταξύ της γραμμικής και της τροπικής άλγεβρας.	40
5.1	Χρόνος εκτέλεσης, τιμές μετρικών ν και ε , ελάχιστος και μέγιστος αριθμός επιζησάντων καταστάσεων κατά μήκος χρονικών πλαισίων για διαφορετικές τιμές της παραμέτρου χαλάρωσης θ για δύο διαφορετικές λέξεις εισόδου. . . .	96
5.2	Τιμές μετρικών ν και ε , ακρίβεια και κάλυψη για διαφορετικές τιμές της παραμέτρου χαλάρωσης θ	104
5.3	Τιμές μετρικών ν και ε , ακρίβεια και κάλυψη για διαφορετικές τιμές της παραμέτρου χαλάρωσης θ	104
5.4	Τιμές μετρικών ν και ε , ακρίβεια και κάλυψη για διαφορετικές τιμές της παραμέτρου χαλάρωσης θ	104
5.5	Τιμές μετρικών ν και ε , ακρίβεια και κάλυψη για διαφορετικές τιμές μεταβλητής χαλάρωσης θ	107
5.6	Τιμές μετρικών ν και ε , ακρίβεια και κάλυψη για διαφορετικές τιμές μεταβλητής χαλάρωσης θ	107
5.7	Τιμές μετρικών ν και ε , ακρίβεια και κάλυψη για διαφορετικές τιμές μεταβλητής χαλάρωσης θ	108
5.8	Μέση τιμή των μεταβλητών χαλάρωσης που υπολογίστηκαν από τον Αλγόριθμο 3, η μέση τιμή του πλήθους των καταστάσεων που εξετάστηκαν, καθώς και η ακρίβεια και η κάλυψη για διαφορετικές τιμές παραμέτρου της αποκλίνουσας κατάστασης.	112
5.9	Μέση τιμή των μεταβλητών χαλάρωσης που υπολογίστηκαν από τον Αλγόριθμο 3, η μέση τιμή του πλήθους των καταστάσεων που εξετάστηκαν, καθώς και η ακρίβεια και η κάλυψη για διαφορετικές τιμές παραμέτρου της αποκλίνουσας κατάστασης.	114
5.10	Μέση τιμή των μεταβλητών χαλάρωσης που υπολογίστηκαν από τον Αλγόριθμο 3, η μέση τιμή του πλήθους των καταστάσεων που εξετάστηκαν, καθώς και η ακρίβεια και η κάλυψη για διαφορετικές τιμές παραμέτρου της αποκλίνουσας κατάστασης.	114
5.11	Μέση τιμή των μεταβλητών χαλάρωσης που υπολογίστηκαν από τον Αλγόριθμο 3, η μέση τιμή του πλήθους των καταστάσεων που εξετάστηκαν, καθώς και η ακρίβεια και η κάλυψη για διαφορετικές τιμές παραμέτρου της αποκλίνουσας κατάστασης.	116

Κεφάλαιο 1

Εισαγωγή

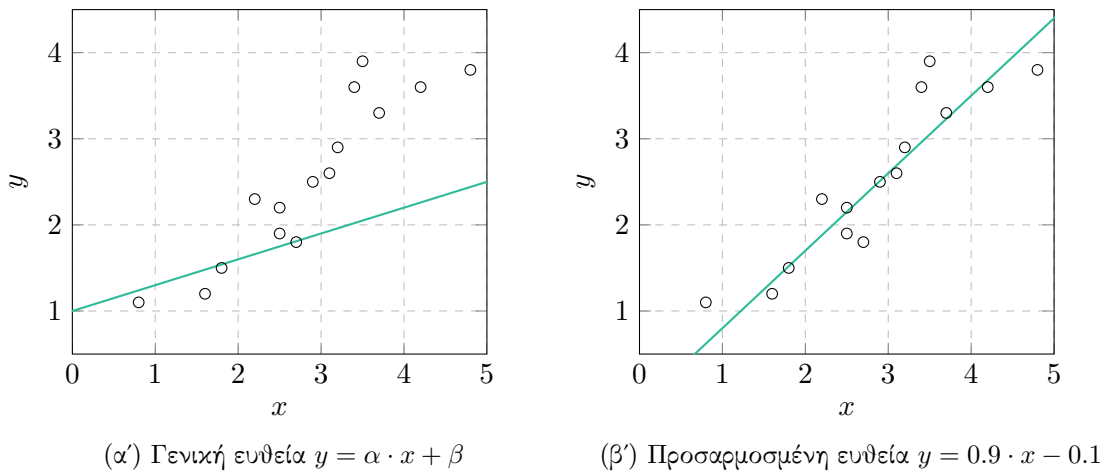
1.1 Μηχανική Μάθηση

Η μηχανική μάθηση (machine learning) αποτελεί ένα από τα πιο θερμά επιστημονικά αντικείμενα του 21^{ου} αιώνα και τα τελευταία χρόνια τόσο το ενδιαφέρον όσο και η έρευνα στο αντικείμενο εξελίσσεται με συνταρακτικούς ρυθμούς. Συχνά, όμως, όταν υπάρχει έξαρση ενδιαφέροντος για έναν επιστημονικό κλάδο, παρατηρείται η φήμη του να προτρέχει της ευρείας κατανόησής του. Συνεπώς, τείνει να υπάρχει μεγάλη σύγχυση γύρω από το τι ακριβώς είναι η μηχανική μάθηση, σε ποιους επιστημονικούς κλάδους βασίζεται για την παραγωγή των αποτελεσμάτων της, ποιες οι, εν τέλει, οι δυνατότητές της και για πιο λόγω αποτελεί ενδιαφέρον αντικείμενο μελέτης, και, κυριότερα, ποιοι ακριβώς είναι οι περιορισμοί που διέπουν το αντικείμενό της. Ως αποτέλεσμα η μηχανική μάθηση αποτελεί συχνά αντικείμενο συζητήσεων και ειδήσεων, χωρίς, όμως, να συνδράμει κάποια ουσιώδη ανάλυση ή κατανόηση του αντικειμένου. Συνεπώς, θα θέλαμε να εκκινήσουμε αυτή τη διπλωματική εργασία με μια σύντομη εισαγωγή στην οποία επιθυμούμε να δώσουμε μια βασική εισαγωγή στις βασικές ιδέες του αντικειμένου, τους στόχους που αποσκοπεί να επιτύχει και τα προβλήματα που επιδιώκει να λύσει η μηχανική μάθηση, καθώς και να κάνουμε μια νύξη στη διαίσθηση των παραπάνω. Ο απώτερος σκοπός της ενότητας είναι να ξεδιαλύνουμε την έννοια της μηχανικής μάθησης στον αναγνώστη, και να καθιστεί εμφανές τι, εν τέλει, είναι και τι δεν είναι δυνατό με τη χρήση της μηχανικής μάθησης.

1.1.1 Βασικές ιδέες

Η βασική ιδέα πάνω στην οποία βασίζεται η μηχανική μάθηση είναι ότι η φύση ακολουθεί κάποια κανονικότητα. Συγκεκριμένα, υποθέτει ότι αρκετά φαινόμενα στον κόσμο έχουν κάποιο ‘απλό’ αίτιο το οποίο διέπει την εξέλιξή τους, και συνεπώς η εύρεση αυτού του αιτίου δύναται, σε ικανοποιητικό βαθμό, να προβλέψει την εξέλιξη του φαινομένου. Σε διαφορετικό πλαίσιο, τα θεμέλια της μηχανικής μάθησης υποθέτουν την ύπαρξη απλών σχέσεων οι οποίες συνδέουν ποσότητες μεταξύ τους. Χαρακτηριστικό παράδειγμα αποτελούν τα ακίνητα. Συγκεκριμένα, η τιμή ενός ακινήτου εξαρτάται, με κάποια συνάρτηση, από την τοποθεσία στην οποία βρίσκεται αυτό το ακίνητο, τα υλικά κατασκευής του, την έκτασή του, κτλ. Αντίστοιχα, οι τιμές των μετοχών, αν και σημαντικά πιο σύνθετες οντότητες, εξαρτώνται σε κάποιο βαθμό από τα προϊόντα τα οποία παράγει η εκάστοτε εταιρία, την εκτίμηση για πιθανές δυνατότητες εξέλιξής της, την εικόνα που έχει διαμορφώσει το καταναλωτικό κοινό για την εταιρία κ.ά.

Έχοντας ως έναυσμα ότι τα φαινόμενα που απαντώνται στον κόσμο εξηγούνται με κάποιες σχέσεις μεταξύ ποσοτήτων, η μηχανική μάθηση αποσκοπεί στην εξαγωγή αυτών των σχέσεων. Στον τομέα της Φυσικής, οι ερευνητές αφιερώνουν μήνες, ίσως χρόνια, στη μελέτη ενός φαινομένου προτείνοντας μοντέλα και καταστρώνοντας επιμελώς σύνθετες μαθηματικές εξισώσεις,



Σχήμα 1.1: Γραμμική υπόθεση για την αλληλεπίδραση μεταξύ δύο ποσοτήτων και προσαρμογή της καμπύλης στα δεδομένα.

στην αναζήτηση κάποιας μοντελοποίησης ικανής να αναπαραστήσει, επαρκώς, την εξέλιξη του φαινομένου. Αντίθετα, η μηχανική μάθηση ακολουθεί μια διαφορετική προσέγγιση για την εξαγωγή των σχέσεων που διέπουν τα αντικείμενα μελέτης της. Συγκεκριμένα, επιχειρεί να εξαγάγει τις σχέσεις αλληλεπίδρασης προσεγγιστικά, αξιοποιώντας διαθέσιμα δεδομένα. Προσπαθεί, δηλαδή, να προσδιορίσει την καμπύλη μεταξύ των ποσοτήτων που εξετάζει η οποία δύναται να επεξηγήσει καλύτερα τα διαθέσιμα δεδομένα για το πρόβλημα, πάντα σύμφωνα με κάποιο κριτήριο. Ο κυριότερος λόγος για τον οποίο η παραπάνω μέθοδος αποτελεί βιώσιμη στρατηγική για την επίλυση προβλημάτων είναι η αφθονία διαθέσιμων δεδομένων στην σύγχρονη εποχή (γεγονός το οποίο ευθύνεται και για την πρόσφατη άνθηση του τομέα). Δεδομένα για κάθε λογής πρόβλημα υπάρχουν διαθέσιμα, και ακόμα και σε περιπτώσεις όπου υπάρχει έλλειψη δεδομένων, η συλλογή τους καθίσταται εξαιρετικά γρήγορη, εύκολη, και οικονομικά εφικτή χάριν του διαδικτύου.

Αξιοποιώντας τα διαθέσιμα δεδομένα, οι ερευνητές καλούνται να κάνουν μια υπόθεση για τον τύπο αλληλεπίδρασης των δεδομένων. Αυτή η αρχική υπόθεση μπορεί να είναι αρκετά αυθαίρετη, αλλά μπορεί επίσης να προέρχεται από κάποια πρώιμη γνώση των ερευνητών σχετικά με τον τομέα εξέτασης, ή να πηγάζει από τη διαίσθησή τους. Είναι δυνατό οι ερευνητές να έχουν προβεί σε κάποια προκαταρκτική μελέτη των δεδομένων και να έχουν εντοπίσει, εποπτικά, τη μορφή της αλληλεπίδρασης των ποσοτήτων. Για παράδειγμα, στο πεδίο της κοστολόγησης ακινήτων, μπορεί να υποθέσει ένας ερευνητής ότι η τιμή του ακινήτου συνδέεται με γραμμικό τρόπο με την συνολική του έκταση. Έχοντας στη διάθεση τους μια αρχική υπόθεση, οι ερευνητές επικαλούνται την μηχανική μάθηση για τον προσδιορισμό των παραμέτρων της υπόθεσης, με βάση τα δεδομένα που έχουν στη διάθεσή τους. Η διαδικασία του προσδιορισμού των παραμέτρων ονομάζεται προσαρμογή (fitting) ή εκπαίδευση του μοντέλου, και αποτελεί τον κορμό της μηχανικής μάθησης. Για παράδειγμα, στην περίπτωση της κοστολόγησης ακινήτων, οι αλγόριθμοι της μηχανικής μάθησης θα αποσκοπούσαν στον εντοπισμό των παραμέτρων της ευθείας (δηλαδή την κλίση και την τομή της) οι οποίες συνδράμουν βέλτιστα με τα δεδομένα, με βάση κάποιο κριτήριο. Στο Σχήμα 1.1 παρουσιάζουμε ένα παράδειγμα προσαρμογής μιας ευθείας. Συγκεκριμένα, γίνεται μια αρχική υπόθεση για την αλληλεπίδραση των ποσοτήτων x και y (για παράδειγμα της έκτασης με την τιμή ενός ακινήτου), και συγκεκριμένα ότι συνδέονται γραμμικά. Έπειτα, στη δεύτερη παράσταση του σχήματος γίνεται η προσαρμογή της ευθείας έτσι ώστε να ελαχιστοποιεί το τετραγωνικό κριτήριο.

Έχοντας κάνει την προκαταρκτική υπόθεση για τα δεδομένα, και έχοντας προσδιορίσει τις παραμέτρους, μέσω των αλγορίθμων της μηχανικής μάθησης, οι ερευνητές καλούνται να αξιολογήσουν την επίδοση του μοντέλου. Συγκεκριμένα, οι ερευνητές επιθυμούν να εξετάσουν κατά πόσο το προκύπτον μοντέλο, με τις συγκεκριμένες παραμέτρους, έχει αντίκτυπο στην πραγματικότητα. Η αξιολόγηση αυτή γίνεται μέσω την εφαρμογή του μοντέλου σε νέα δεδομένα του ίδιου αντικειμένου εφαρμογής, τα οποία δεν χρησιμοποιήθηκαν κατά τη διαδικασία εκπαίδευσης, έτσι ώστε να ελεγχθεί εάν οι προβλέψεις που παράγει το μοντέλο είναι ικανοποιητικά κοντά στις παρατηρούμενες τιμές. Εάν κάτι τέτοιο δεν είναι αληθές, γίνεται αναθεώρηση του μοντέλου, δηλαδή οι ερευνητές κάνουν μια διαφορετική υπόθεση για την σχέση μεταξύ των εξεταζόμενων ποσοτήτων, και επανεκπαιδεύουν το μοντέλο χρησιμοποιώντας τα διαθέσιμα δεδομένα. Κατανοούμε, λοιπόν, ότι η μηχανική μάθηση αποτελεί μια επαναληπτική διαδικασία μετά την ολοκλήρωση της οποίας η διαδικασία ξεκινάει εκ νέου, εάν αυτό κριθεί απαραίτητο. Τα τρία κύρια τμήματα της διαδικασίας είναι:

- Η προκαταρκτική υπόθεση των ερευνητών για την φύση της αλληλεπίδρασης μεταξύ των ποσοτήτων υπό μελέτη.
- Η προσαρμογή του προτεινόμενου μοντέλου στα διαθέσιμα δεδομένα με απώτερο σκοπό τον προσδιορισμό των υπερπαραμέτρων του.
- Η αξιολόγηση της επίδοσης του μοντέλου σε νέα δεδομένα, έτσι ώστε να εξεταστεί εάν η συγκεκριμένη μοντελοποίηση δύναται να αναπαραστήσει επαρκώς την πραγματική σχέση μεταξύ των ποσοτήτων.

1.1.2 Κριτήρια ελαχιστοποίησης

Στην Ενότητα 1.1.1 αναφέραμε ότι οι αλγόριθμοι μηχανικής μάθησης προσπαθούν να προσδιορίσουν τις υπερπαραμέτρους οι οποίες οδηγούν στην ελαχιστοποίηση κάποιων κριτηρίων, οδηγώντας στην βέλτιστη παραμετροποίηση του μοντέλου. Ο κύριος λόγος για τον οποίο χρησιμοποιούνται διαφορετικά κριτήρια ελαχιστοποίησης είναι ότι, για το ίδιο πρόβλημα, μπορεί να υπάρχουν πολλαπλοί ορισμοί βελτιστότητας. Για παράδειγμα, ας εξετάσουμε την περίπτωση δύο ταξιδιωτών οι οποίοι σχεδιάζουν κάποιο πιθανό ταξίδι και προσπαθούν να χαράξουν την πορεία που θα ακολουθήσουν. Οι ταξιδιώτες μπορεί να επιθυμούν να ελαχιστοποιήσουν το συνολικό κόστος του ταξιδιού. Διαφορετικά, μπορεί να επιθυμούν να μεγιστοποιήσουν τον αριθμό των μνημείων που θα επισκεφθούν, ή να ελαχιστοποιήσουν τους χρόνους μετακίνησης. Παρατηρούμε, λοιπόν, ότι για το πρόβλημα της κατάστρωσης ενός ταξιδιού υπάρχουν διαφορετικά πιθανά κριτήρια βελτιστοποίησης, ανάλογα με τις ανάγκες του εκάστοτε χρήστη. Κρίνουμε, λοιπόν, σκόπιμο σε αυτό το σημείο να αφιερώσουμε μια σύντομη ενότητα στην μελέτη μερικών πολύ συχνά χρησιμοποιούμενων κριτηρίων ελαχιστοποίησης, καθώς και επίσης να εξηγήσουμε συνοπτικά τις προσφορές του καθ' ενός (και άρα να εξηγήσουμε υπό ποιες συνθήκες θα επέλεγαν τη χρήση του καθ' ενός οι ερευνητές). Η επιλογή μας για την ανάλυση πιθανών κριτηρίων ελαχιστοποίησης βασίζεται στο γεγονός ότι μεγάλο μέρος της διπλωματικής εργασίας, καθώς και τα αποτελέσματα αυτής, αφιερώνεται στην προσπάθεια ελαχιστοποίησης κριτηρίων, ή εύρεσης βέλτιστων λύσεων υπό συνθήκες. Συνεπώς, θεωρούμε επιτακτική την παρουσίαση μιας τέτοιας ενότητας.

Τα κριτήρια ελαχιστοποίησης, τα οποία έχουν τις γενικότερες ρίζες τους στο πεδίο της Βελτιστοποίησης ([3], [8], [45]), αποτελούν μαθηματικές συναρτήσεις οι οποίες ορίζονται πάνω στα μεγέθη ενός προβλήματος μηχανικής μάθησης. Συγκεκριμένα, συχνά έχουν ως είσοδο την καταγεγραμμένη πραγματική τιμή (ground truth) που προσπαθεί να προσομοιώσει κάποιος αλγόριθμος μηχανικής μάθησης (για παράδειγμα, την τιμή στην οποία πωλήθηκε ένα ακίνητο) καθώς και την εκτίμηση αυτής της τιμής που παράγει ο αλγόριθμος. Ο υπολογισμός που

εκτελούν αυτές οι συναρτήσεις είναι μια έκφραση του πόσο απέχουν οι δύο είσοδοι, σύμφωνα με το εκάστοτε κριτήριο. Η γενική μορφή αυτών των συναρτήσεων είναι:

$$L(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}) = \sum_i f(\theta_i, \bar{\theta}_i) \quad (1.1)$$

όπου f είναι η συνάρτηση η οποία χαρακτηρίζει το κριτήριο βελτιστοποίησης. Όπως αναφέραμε, οι συναρτήσεις υπολογίζουν τον βαθμό, στο πλαίσιο του κριτηρίου, στον οποίο οι προβλέψεις του εκάστοτε μοντέλου μηχανικής μάθησης συνάδουν με τα διαθέσιμα δεδομένα. Η συνάρτηση f καθορίζει την ερμηνεία της σύναξης στο εκάστοτε πλαίσιο.

Εάν και υπάρχει πληθώρα κριτηρίων ελαχιστοποίησης για περίσσεια προβλημάτων, δύο κριτήρια απαντώνται με μεγαλύτερη συχνότητα στο επιστημονικό πεδίο της μηχανικής μάθησης. Τα δύο αυτά κριτήρια αποσκοπούν στην ελαχιστοποίηση καίριων μεγεθών για την μηχανική μάθηση. Συγκεκριμένα, τα προκείμενα συναρτησιακά αποσκοπούν:

- στην ελαχιστοποίηση του τετραγωνικού λάθους. Το λάθος ορίζεται ως η διαφορά μεταξύ των καταγεγραμμένων τιμών και των εκτιμήσεων που παράγει ο αλγόριθμος μηχανικής μάθησης, και η τετραγωνοποίηση της ποσότητας αυτής έχει μερικές πολύ θεμιτές ιδιότητες (όπως παραγωγισιμότητα, συνέχεια, συμμετρία, κλπ). Συνεπώς, είναι εμφανές για ποιους λόγους ένα τέτοιο κριτήριο είναι απαραίτητο: εάν ένα μοντέλο μηχανικής μάθησης παράγει προβλέψεις οι οποίες αδυνατούν να μοντελοποιήσουν επαρκώς τα δεδομένα, τότε το μοντέλο αυτό, για το συγκεκριμένο πρόβλημα, δεν είναι χρήσιμο. Το κριτήριο το οποίο, μαθηματικά, περιγράφει το τετραγωνικό λάθος (γνωστού και ως L_2 νόρμα) δίνεται από την σχέση:

$$L_2(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}) = \sum_i (\theta_i - \bar{\theta}_i)^2 \quad (1.2)$$

- στην ελαχιστοποίηση της διαστατικότητας του μοντέλου. Η ελαχιστοποίηση αυτή αποτελεί μια ιδιαίτερα ενδιαφέρουσα μορφή ελαχιστοποίησης. Τα αραιά μοντέλα αναγνωρίστηκαν, δηλαδή μοντέλα με χαμηλή διαστατικότητα, είχαν αναγνωριστεί από τον 13^ο αιώνα μ.Χ. από τον William of Ockham¹ (περί το 1287–1347 μ.Χ.). Ουσιαστικά, ο Ockham υποστήριζε ότι όταν υπάρχουν πολλαπλές εξηγήσεις για ένα φαινόμενο, η πιο απλή είναι αυτή με την μεγαλύτερη πιθανότητα. Για παράδειγμα, εάν εισέλθει ένας φοιτητής μέσα σε ένα αμφιθέατρο τον μήνα Δεκέμβρη και είναι βρεγμένος, είναι πολύ πιο πιθανό να βρέχει έξω και να μην είχε ομπρέλα (το οποίο αποτελεί μια απλή εξήγηση), παρά να μπήκε σε πισίνα και να κολύπησε με τα ρούχα πριν την διάλεξη (το οποίο αποτελεί μια σημαντικά πιο σύνθετη εξήγηση). Ο αναγνώστης, όμως, πρέπει να θυμάται ότι υπάρχουν προβλήματα για τα οποία δεν επιδέχονται απλές εξηγήσεις. Για παράδειγμα, στην αναγνώριση προσώπων τα μοντέλα τα οποία έχουν εξαιρετική επίδοση είναι αρκετά σύνθετα, και αποτελούνται από πολλά στάδια συνελικτικών δικτύων. Σε τέτοια προβλήματα, εάν είναι κρίσιμη για την εφαρμογή η αναγνώριση σε πραγματικό χρόνο (όπως για παράδειγμα σε συστήματα καθοδήγησης ρομπότ), συχνά πρέπει να θυσιάσουμε την ακρίβεια στο βωμό της γρηγορότερης αναγνώρισης. Αυτό επιτυγχάνεται με την χρήση απλούστερων μοντέλων, τα οποία είναι μεν υποδεέστερα σε επιδόσεις, αλλά είναι δε σημαντικά γρηγορότερα. Τέλος, οι αραιές αναπαραστάσεις λειτουργούν και ως ένα αντίβαρο στην ελαχιστοποίηση του λάθους. Οι αραιές αναπαραστάσεις δεν μπορούν να χρησιμοποιηθούν αυτόνομες, καθώς κάτι τέτοιο θα οδηγούσε σε ένα κενό μοντέλο, καθώς αυτό ελαχιστοποιεί πλήρως το κριτήριο και δεν υπάρχει κάποιος άλλος περιορισμός.

¹Ευράφι του Ockham: 'Υπό την ύπαρξη πολλών πιθανών εξηγήσεων για ένα φαινόμενο, η απλούστερη τείνει να είναι, συνήθως, η σωστή.'

Λειτουργούν ως αντίβαρο στην ελαχιστοποίηση του λάθους με τον εξής τρόπο: εάν αφαιρεθεί πλήρως ελεύθερο ένα μοντέλο ελαχιστοποίησης του σφάλματος, τότε θα θεωρεί όλο και πιο σύνθετα σενάρια, τα οποία θα εξηγούν ακριβώς την παραγωγή των δεδομένων, δηλαδή η προτεινόμενη καμπύλη θα διέρχεται από όλα τα σημεία δεδομένων. Όμως, συχνά τα δεδομένα έχουν ανακρίβειες και διακυμάνσεις. Για παράδειγμα, ένα σπίτι αντικειμενικής αξίας 250,000 ευρώ έχει μεγάλη πιθανότητα να πωληθεί, για πληθώρα λόγων, από 150,000 ευρώ έως 350,000 ευρώ. Η ελαχιστοποίηση της διαστατηρότητας βάζει ένα ‘φρένο’ στο κατά πόσο επιδιώκουμε την απόλυτη ελαχιστοποίηση του λάθους. Η ελαχιστοποίηση γίνεται μέσω της L_0 νόρμας, δηλαδή των πόσων μη μηδενικών παραμέτρων επιδέχεται το μοντέλο. Όμως, κάτι τέτοιο είναι μαθηματικά δύστροπο, καθώς δεν ορίζεται με κλειστό μαθηματικό τύπο. Εκτός αυτού, δεν ορίζεται παράγωγος. Αντ’ αυτής χρησιμοποιείται συνήθως η L_1 νόρμα:

$$L_1(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}) = \sum_i |\theta_i - \bar{\theta}_i| \quad (1.3)$$

Ο λόγος για τον οποίο μας ενδιαφέρει το παραπάνω κριτήριο ελαχιστοποίησης είναι επειδή οδηγεί σε κάτι που ονομάζουμε *αραιές λύσεις*. Οι αραιές λύσεις μηδενίζουν κάποιες από τις παραμέτρους των μοντέλων, οδηγώντας, συνεπώς, σε απλούστερα μοντέλα, και άρα σε γρηγορότερους χρόνους εκτέλεσης.

1.1.3 Διεργασίες μηχανικής μάθησης

Η βασική ιδέα της μηχανικής μάθησης είναι ότι από έναν όγκο δεδομένων θα εξαχθεί, αυτόματα, κάποια πληροφορία. Αυτό συνήθως γίνεται με την επιλογή ενός μοντέλου, για παράδειγμα ότι υπάρχει μια γραμμική σχέση μεταξύ των χρημάτων και της ευτυχίας. Έπειτα, χρησιμοποιείται ο όγκος δεδομένων ώστε να υπολογιστεί ποια ακριβώς είναι αυτή η σχέση, και να αξιολογηθεί κατά πόσο η υπόθεση για το μοντέλο επαφίεται της πραγματικότητας. Αναφορικά με τους στόχους, η μηχανική μάθηση αποσκοπεί στην αντιμετώπιση δύο κλασικών δοκιμασιών. Οι δύο βασικές κατηγορίες προβλημάτων τα οποία καλείται να επιλύσει είναι:

- **Επιβλεπόμενη μάθηση.** Στα προβλήματα επιβλεπόμενης μάθησης τα διαθέσιμα δεδομένα έχουν κάποια επισήμειωση, την οποία οι αλγόριθμοι καλούνται να προβλέψουν χρησιμοποιώντας τα επιμέρους χαρακτηριστικά των δεδομένων. Μερικά χαρακτηριστικά παραδείγματα δεδομένων τα οποία απαντώνται με μεγάλη συχνότητα στον κλάδο, τα οποία επιδέχονται λύση επιβλεπόμενης μάθησης, είναι:

- δεδομένα με χαρακτηριστικά σπιτιών (όπως η περιοχή, το μέγεθος, ο αριθμός υπνοδωματίων, η χρονολογία κατασκευής) μαζί με το αντίστοιχο κόστος, όπου ο αλγόριθμος καλείται να προβλέψει την τιμή σπιτιών με βάση τα χαρακτηριστικά τους. Τα προβλήματα αυτά ονομάζονται προβλήματα *παλινδρόμησης* (regression) όπου ο αλγόριθμος καλείται να επεξεργαστεί αριθμητικά τα διαθέσιμα χαρακτηριστικά, αποδίδοντας τους βάρη, με σκοπό να προβλέψει μια συγκεκριμένη τιμή-στόχο. Συγκεκριμένα, σε αυτά τα προβλήματα η πρόβλεψη \bar{y} του αλγορίθμου δίνεται από την εξίσωση της παλινδρόμησης:

$$\bar{y} = w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_n \cdot x_n = \mathbf{w}^T \mathbf{x} \quad (1.4)$$

όπου \mathbf{x} είναι το διάνυσμα των χαρακτηριστικών για το εκάστοτε εξεταζόμενο σπίτι, και \mathbf{w} είναι το διάνυσμα βαρών που καλούνται να προσδιορίσουν οι αλγόριθμοι μάθησης για να προβλέψουν το επιθυμητό μέγεθος.

- εικόνες αντικειμένων μαζί με την επισημείωση περί τον χαρακτήρα του αντικείμενου, όπου ο αλγόριθμος καλείται να μπορεί να αναγνωρίσει πιο αντικείμενο αναπαρίσταται σε μια εικόνα. Τα προβλήματα αυτά ονομάζονται προβλήματα *αναγνώρισης* (classification) και ο αλγόριθμος καλείται να προβλέψει μια επισημείωση. Συχνά, τα προβλήματα αναγνώρισης μοντελοποιούνται ως πλήθος προβλημάτων παλινδρόμησης, όπου ως επισημείωση επιλέγεται αυτή στην οποία ανατίθεται η μεγαλύτερη πιθανότητα.
- δεδομένα από το χρηματιστήριο μαζί με τις τιμές των μετοχών, όπου ο αλγόριθμος καλείται να προβλέψει μελλοντικές τιμές των μετοχών. Ο λόγος που η συγκεκριμένη εφαρμογή παρίσταται ξεχωριστά από την πρώτη, παρά τις εμφανείς διαφορές τους, είναι ότι τα χρηματιστηριακά δεδομένα παρουσιάζουν το ενδιαφέρον φαινόμενο της περιοδικότητας. Συγκεκριμένα, ενώ για την μεσιτική εφαρμογή η σχέση μεταξύ των κατοικιών και τις τιμές τους είναι σχετικά σταθερή, η σχέση των μετοχών επηρεάζεται από περιοδικούς παράγοντες (νέα, εποχή, ρευστότητα). Επιπλέον, συχνά σε αυτές τις εφαρμογές ζητείται η συνεχής πρόβλεψη τιμών και η εξέλιξή τους στο χρόνο.

Δύο ιδιαίτερες περιπτώσεις οι οποίες χρήζουν μνείας είναι:

- **Active learning:** Σε αυτή την ιδιαίτερη περίπτωση, ο αλγόριθμος δεν έχει διαθέσιμες επισημειώσεις, αλλά μπορεί να αποκτήσει επιλεκτικά κάποιες. Έτσι, σε αυτή τη δοκιμασία ο στόχος είναι να ελαχιστοποιηθεί το κόστος το οποίο θα υποφέρει ο μηχανικός κατά τη διαδικασία συλλογής δεδομένων, πετυχαίνοντας, παράλληλα, το βέλτιστο δυνατό αποτέλεσμα.
 - **Reinforcement learning:** Σε αυτή την περίπτωση οι ερευνητές προσπαθούν να προσομοιάσουν την διαδικασία μάθησης που ακολουθούν οι τα νεογέννητα μωρά. Συγκεκριμένα, αναπτύχθηκε επειδή τα μοντέλα τα οποία προτεινόταν για την επίλυση προβλημάτων γινόταν ολόένα και πιο περίπλοκα, χωρίς, όμως, να έχει υπάρξει κάποια εξήγηση για τις επιδόσεις των μοντέλων. Έτσι, το reinforcement learning αποτέλεσε μια προσπάθεια να δημιουργηθεί μια κατηγορία μοντέλων τα οποία να έχουν πολύ απλή δομή. Αυτό επιτυγχάνεται με ένα ιδιαίτερο περιορισμό στην αξιολόγησή τους: η μόνη ανατροφοδότηση που επιδέχονται είναι μόνο εάν η πρόβλεψή τους ήταν ορθή ή λανθασμένη. Έτσι, προσομοιάζεται η διαδικασία μάθησης των νεογέννητων μωρών, που αξιολογούν τις πράξεις τους μονάχα ως προς έναν άξονα: αν η πράξη εισέπραξε επιβράβευση ή επίπληξη.
- **Μη-επιβλεπόμενη μάθηση.** Στα προβλήματα μη-επιβλεπόμενης μάθησης τα διαθέσιμα δεδομένα έχουν διαφορετική δομή. Συγκεκριμένα, υπάρχει κάποιος όγκος δεδομένων με τα αντίστοιχα, αλλά απουσιάζει παντελώς η οποιαδήποτε επισημείωση. Ο σκοπός σε αυτή τη δοκιμασία είναι να εξεταστούν οι συσχετίσεις μεταξύ των διαθέσιμων δεδομένων και να εντοπιστεί κάποια ιεραρχική δομή σε έναν φαινομενικά αχανή όγκο δεδομένων. Μερικά χαρακτηριστικά προβλήματα είναι μη-επιβλεπόμενης μάθησης είναι:
- *Συσταδοποίηση* (clustering) των δεδομένων. Καθημερινά απαντώμενο παράδειγμα αποτελεί η ομαδοποίηση άρθρων ανάλογα με θεματικές ενότητες. Αναλύοντας το μήκος των άρθρων, το ιστόγραμμα των λέξεων που χρησιμοποιούνται στο καθ' ένα, και άλλα στατιστικά μεγέθη που υπολογίζονται ανάλογα με τις απαντώμενες λέξεις, οι αλγόριθμοι προσπαθούν να ομαδοποιήσουν άρθρα τα οποία έχουν κοινές θεματικές ενότητες. Μέσω της ομαδοποίησης αυτής επιθυμούμε να προτείνουμε επιπλέον άρθρα σε έναν αναγνώστη τα οποία να σχετίζονται με τη θεματική ενότητα για την οποία ενδιαφέρεται.

- Αναγνώριση ανωμαλιών. Το πιο καθημερινό παράδειγμα αποτελεί η προσπάθεια των τραπεζών για αυτόματη αναγνώριση απατηλών χρεώσεων. Συγκεκριμένα, αξιοποιώντας το πρώιμο ιστορικό συναλλαγών του χρήστη ο αλγόριθμος μηχανικής μάθησης καλείται να εκφανθεί εάν μια νέα χρέωση είναι γνήσια χρέωση του δικαιούχου του λογαριασμού ή όχι, στην οποία περίπτωση γνωστοποιεί την πρόβλεψή του και απαιτείται επιβεβαίωση για την εκτέλεση της συναλλαγής. Αρχικά, ο αναγνώστης μπορεί να θεωρήσει ότι το εν λόγω πρόβλημα μπορεί να επιλυθεί επικαλώντας επιβλεπόμενη μάθηση, θεωρώντας ως επισημειώσεις την πληροφορία για τις έγκυρες συναλλαγές. Όμως, το συγκεκριμένο πρόβλημα διαφέρει στη φύση, καθώς θα θέλαμε να είμαστε σε θέση να αποφανθούμε για τις απατηλές χρεώσεις χωρίς να έχουμε κανένα πρώιμο δεδομένο απατηλής χρέωσης για τον συγκεκριμένο δικαιούχο. Έτσι, το πρόβλημα δεν έγκειται στον εντοπισμό της ορθής κατηγοριοποίησης για μια νέα συναλλαγή, αλλά στην απόφαση εάν η νέα συναλλαγή διαφέρει σημαντικά από το μοτίβο του πρώιμου ιστορικού του εκάστοτε δικαιούχου.
- Πρόταση προϊόντων. Σε αυτά τα προβλήματα υπάρχει διαθέσιμος όγκος δεδομένων για αγορές και βαθμολογίες προϊόντων από χρήστες, και ο σκοπός είναι η πρόταση νέων προϊόντων τα οποία κρίνει ο αλγόριθμος μηχανικής μάθησης ότι θα ενδιαφέρουν το χρήστη. Η ιδιαιτερότητα του προβλήματος έγκειται ότι για τις προτάσεις που κάνει ο αλγόριθμος σε κάθε διαφορετικό χρήστη, αξιοποιεί όλη την διαθέσιμη πληροφορία από το σύνολο των χρηστών. Συνεπώς, το πρόβλημα έγκειται στην εξαγωγή σχέσεων χαμηλού επιπέδου μεταξύ των χρηστών και των προϊόντων. Για παράδειγμα, εάν ένας χρήστης αγοράζει πολλά αναψυκτικά μιας συγκεκριμένης εταιρίας, είναι πιθανό να του αρέσουν γενικά τα αναψυκτικά, επομένως ο αλγόριθμος θα μπορούσε να του προτείνει άλλα αναψυκτικά, τα οποία οι υπόλοιποι χρήστες, που έχει στη διάθεσή του ο αλγόριθμος, έχουν προτιμήσει.

1.2 Προβλήματα μηχανικής μάθησης

Η μηχανική μάθηση, κατά κοινή ομολογία, εξελίσσεται με εκρηκτικούς ρυθμούς τα τελευταία χρόνια. Το ενδιαφέρον των ερευνητών σε παγκόσμιο επίπεδο για συστήματα τα οποία επιλύουν προβλήματα με επίδοση άμεσα συγκρίσιμη με αυτή των ανθρώπων, συνδυασμένο με την αφθονία των διαθέσιμων δεδομένων και της υπολογιστικής ισχύς, έχουν οδηγήσει σε πραγματικά αξιοθαύμαστα αποτελέσματα. Συγκεκριμένα, σε ορισμένα προβλήματα, τα συστήματα που έχουν αναπτυχθεί όχι μόνο προσομοιάζουν την επίδοση των ανθρώπινων ειδικών, αλλά ξεπερνάνε την επίδοσή τους.

Τα παραπάνω έχουν οδηγήσει στην άνθιση του τομέα, καθιστώντας τον έναν από τους ενεργούς ερευνητικούς τομείς, με πληθώρα αποτελεσμάτων. Παρ' όλα αυτά, η ραγδαία ανάπτυξη έχει το τίμημά της. Ο ζήλος με τον οποίο συχνά οι ερευνητές αναπτύσσουν τα συστήματα τους παρασύρει στην συνεχή και άβουλη ανάπτυξη συστημάτων, χωρίς αυτή η ανάπτυξη να βασίζεται σε εκτεταμένη μαθηματική σκέψη και διαισθητικές παρατηρήσεις. Ο κύριος τρόπος με τον οποίο αντιμετωπίζονται τα τέλματα στις επιδόσεις των αλγορίθμων μηχανικής μάθησης από τους μηχανικούς είναι με την υπόθεση όλο και πιο σύνθετων μοντέλων και όλο και μεγαλύτερων αρχιτεκτονικών. Αυτό έχει ως αποτέλεσμα η έρευνα να μην έχει κάποια συγκεκριμένη κατεύθυνση. Αντίθετα, οι ερευνητές συνθέτουν και δοκιμάζουν ακατάπαυστα μοντέλα, ελπίζοντας ότι αυτά θα οδηγήσουν στην πολυπόθητη βελτίωση της απόδοσής τους. Το πρόβλημα αυτής της προσέγγισης έγκειται στο ότι, ανεξάρτητα από το εάν τα μοντέλα επιτυγχάνουν την βελτίωση της απόδοσης ή όχι, οι ερευνητές αδυνατούν να κατανοήσουν ποια ακριβώς επιλογή τους επέφερε αυτή τη βελτίωση, και, κυριότερα, γιατί. Έτσι, οι ερευνητές δεν διαθέτουν κάποια πυξίδα στην έρευνα τους: δεν είναι σε θέση να ερμηνεύσουν, ουσια-

δώς, τα αποτελέσματα τους, και μετέπειτα να κινηθούν, εφοδιασμένοι με την πληροφορία της απόδοσης του προηγούμενου μοντέλου και τα ερωτήματα τα οποία απαντήθηκαν από τον πειραματισμό τους, προς τη βέλτιστη κατεύθυνση (όπως ακριβώς κάνουν τα συστήματα τα οποία εκπαιδεύουν).

Υποστηρίζουμε ότι η λύση στα παραπάνω προβλήματα είναι η μαθηματική μοντελοποίηση των αλγορίθμων της μηχανικής μάθησης. Η μηχανική μάθηση αποτελεί πλέον ξεχωριστό ερευνητικό τομέα, και συγκεντρώνει ερευνητές από ποικιλία υποβάρων από πληθώρα ερευνητικών τομέων. Αν και αυτό το γεγονός είναι εξαιρετικά θετικό, καθώς οδηγεί στην εισαγωγή νέων ιδεών στην μηχανική μάθηση και, παράλληλα, βοηθάει στην ανάπτυξη και την επέκτασή της, οδηγεί επίσης και σε αρκετά προβλήματα. Όπως αναφέραμε προηγουμένως, η ανάπτυξη νέων αλγορίθμων γίνεται τυχαία, χωρίς να υπάρχει ουσιώδης κατανόηση του προβλήματος, αλλά ούτε και της λειτουργίας των υπαρχόντων αλγορίθμων. Έτσι, η επίδοση σε κάποιο συγκεκριμένο πρόβλημα είναι συγκυριακή και αποτέλεσμα τύχης, αλλά το κυριότερο μελανό σημείο της προσέγγισης έγκειται στο ότι δεν αποκτώνται διαισθήσεις και πορίσματα για τις ιδιότητες των αλγορίθμων, καθώς και κατανόηση αναφορικά με τις συνθήκες που θα οδηγούσαν στην βέλτιστη αξιοποίηση των αναπτυσσόμενων αλγορίθμων. Θεωρούμε ότι μέσω της ώριμης ανάλυσης των μεθόδων και των μοντέλων που καλούνται να επιλύσουν τα προβλήματα ενδιαφέροντος δύναται να οδηγηθούμε σε ουσιώδη και βαθιά κατανόηση των ιδιοσυγκρασιών τους. Κάτι τέτοιο επιτρέπει, ακόμα και σε περιπτώσεις όπου οι διαστάσεις των δεδομένων είναι αβυσσαλέες, να μπορούν οι ερευνητές να εξάγουν ουσιώδη και καίρια συμπεράσματα, οδηγώντας, εν τέλει, ορθολογικά την έρευνα πάνω στη μηχανική μάθηση σε νέα επίπεδα.

1.3 Στόχος της διπλωματικής εργασίας

Ο στόχος της συγκεκριμένης εργασίας είναι να προσπαθήσει να καλλιεργήσει τη διαίσθηση σε μερικά προβλήματα της μηχανικής μάθησης. Να προσπαθήσει, δηλαδή, μέσω της αυστηρής μαθηματικής τους μοντελοποίησης να εξερευνήσει σε βάθος τα προβλήματα, ανακαλύπτοντας, έτσι, μια δομή σε αυτά και οδηγώντας στην πιο ολοκληρωμένη κατανόησή τους. Έχοντας μια πιο ώριμη κατανόηση του προβλήματος, αλλά και επιπλέον εργαλεία στη διάθεσή τους, οι ερευνητές είναι πλέον σε θέση να κάνουν πιο καίριες και ουσιώδης ερωτήσεις αναφορικά με τη λειτουργία και τη χρησιμότητα του εκάστοτε μοντέλου. Έτσι, η βελτίωσή τους, ευελπιστούμε, να γίνει με έναν πιο συστηματικό και μεθοδικό τρόπο, με γνώμονα πάντα τον μαθηματικό κανόνα.

Συγκεκριμένα, στο πλαίσιο της διπλωματικής εργασίας θα αξιοποιήσουμε το ερευνητικό πεδίο της τροπικής άλγεβρας και της τροπικής γεωμετρίας για να μοντελοποιήσουμε αλγορίθμους της μηχανικής μάθησης πάνω σε γράφους. Θα εφαρμόσουμε τις ιδέες των παραπάνω πεδίων σε αλγορίθμους που δρουν πάνω σε επεκτάσεις των automata από το πεδίο της Θεωρητικής Επιστήμης των Υπολογιστών, οι οποίες απαντώνται συχνά στην αναγνώριση φωνής και στην επεξεργασία φυσικής γλώσσας, αλλά και στον τομέα των τηλεπικοινωνιών. Σε ευρύτερο πλαίσιο, αυτοί οι αλγόριθμοι αποτελούν κλασικούς αλγορίθμους της Επιστήμης των Υπολογιστών, και αφορούν την βελτιστοποίηση των δομών δεδομένων που χρησιμοποιούνται στις προαναφερθείσες εφαρμογές, ή στην εύρεση βέλτιστων λύσεων. Θα μοντελοποιήσουμε τους εν λόγω αλγορίθμους στην τροπική άλγεβρα, και έπειτα θα γίνει προσπάθεια, αξιοποιώντας την τροπική γεωμετρία, να αναλύσουμε την δομή τους και να εξάγουμε συμπεράσματα για την λειτουργία τους. Συγκεκριμένα, η διάρθρωση της υπόλοιπης διπλωματικής εργασίας θα ακολουθεί την εξής δομή:

- Στο Κεφάλαιο 2 γίνεται παρουσίαση των εισαγωγικών εννοιών της τροπικής άλγεβρας και της τροπικής γεωμετρίας και υπογραμμίζονται οι διαφορές τους με τα γραμμικά

τους ανάλογα. Παράλληλα, παρουσιάζονται μερικά παραδείγματα για την εξοικείωση του αναγνώστη, αλλά και για την διαβεβαίωσή του για την χρησιμότητα της τροπικής μοντελοποίησης σε πληθώρα προβλημάτων.

- Στο Κεφάλαιο 3 γίνεται παρουσίαση των βασικών ιδεών, καθώς και της δομής ποικιλίας γραφικών μοντέλων. Τα γραφικά μοντέλα αποτελούν βασικά δομικά στοιχεία της Επιστήμης των Υπολογιστών, και θα αποβούν ιδιαίτερα σημαντικά για την ανάλυσή που θα ακολουθήσει στο υπόλοιπο της διπλωματικής εργασίας.
- Στο Κεφάλαιο 4 γίνεται η εισαγωγή των αλγορίθμων που δρουν πάνω στα γραφικά μοντέλα που εισήγαμε στο Κεφάλαιο 3, καθώς και η ανάλυσή τους στο πεδίο της τροπικής άλγεβρας. Έπειτα, προβαίνουμε και στην ανάλυση των γεωμετρικών ιδιοτήτων, μέσω της τροπικής γεωμετρίας, ενός υποσυνόλου των αλγορίθμων που εισήγαμε.
- Στο Κεφάλαιο 5 γίνεται παρουσίαση των πειραματικών αποτελεσμάτων που εκτελέστηκαν. Συγκεκριμένα, έγινε εκτενής πειραματισμός με την εφαρμογή των ιδεών που αναπτύξαμε στο Κεφάλαιο 4, τόσο σε κάποια βάση δεδομένων, αλλά και μέσω προσομοιώσεων. Επιπλέον, γίνεται ανάπτυξη ενός αλγορίθμου για την αυτόματη προσαρμογή παραμέτρων σε ένα πρόβλημα εύρεσης βέλτιστων μονοπατιών.
- Τέλος, στο Κεφάλαιο 6 ολοκληρώνουμε τη διπλωματική εργασία συνοψίζοντας τις συνεισφορές της και συζητώντας συνοπτικά μελλοντικές ερευνητικές πορείες και κατευθύνσεις που μπορεί να ακολουθήσει η επέκταση των ιδεών της εργασίας.

1.4 Συμβολισμός

Θα αφιερώσουμε την ακόλουθη ενότητα σε μια σύντομη ανάλυση του συμβολισμού που θα χρησιμοποιήσουμε καθ' όλη την διάρκεια του κειμένου. Αρχικά, αναφέρεται ότι θα ακολουθήσουμε τη σύμβαση των διανυσμάτων στηλών, δηλαδή ένα n -διάστατο διάνυσμα έχει διαστάσεις $n \times 1$, και κάθε διάνυσμα για το οποίο γίνεται αναφορά στο κείμενο θα θεωρείται πως είναι σε αυτή τη μορφή.

Επιπλέον, θα χρησιμοποιήσουμε το σύμβολο \mathbb{R} για να αναφερθούμε στη γραμμή των πραγματικών αριθμών, $(-\infty, +\infty)$. Ορίζουμε τα δύο επεκτεταμένα σύνολα των πραγματικών αριθμών ως εξής:

- $\mathbb{R}_{\min} := \mathbb{R} \cup \{+\infty\}$, δηλαδή το σύνολο \mathbb{R}_{\min} αποτελείται από την πραγματική ευθεία εμπλουτισμένη με ένα στοιχείο για την μοντελοποίηση του απείρου.
- $\mathbb{R}_{\max} := \mathbb{R} \cup \{-\infty\}$, δηλαδή το σύνολο \mathbb{R}_{\max} αποτελείται από την πραγματική ευθεία εμπλουτισμένη με ένα στοιχείο για την μοντελοποίηση του μείον άπειρο.

Επιπλέον, διευκρινίζεται ότι θα συμβολίζουμε:

- όλα τα βαθμωτά μεγέθη με πεζά λατινικά ή ελληνικά γράμματα.
- όλα τα διανύσματα με έντονα πεζά λατινικά ή ελληνικά γράμματα.
- όλους τους πίνακες με έντονα κεφαλαία λατινικά ή ελληνικά γράμματα.

Για παράδειγμα, τα παρακάτω αποτελούν ένα βαθμωτό μέγεθος, ένα διάνυσμα διαστάσεων 3×1 , και έναν πίνακα διαστάσεων 2×3 , αντίστοιχα:

$$\alpha = 5, \quad \mathbf{w} = \begin{bmatrix} 0 \\ 8 \\ 3 \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} 3 & 0 & 1 \\ 5 & -2 & 4 \end{bmatrix}$$

Όταν αναφερόμαστε σε συγκεκριμένα στοιχεία πινάκων ή διανυσμάτων τότε χρησιμοποιούμε τα μη έντονα σύμβολα με τους αντίστοιχους δείκτες. Για παράδειγμα, το πρώτο στοιχείο του παραπάνω διανύσματος θα συμβολίζεται με w_1 .

Κεφάλαιο 2

Τροπική Άλγεβρα και Γεωμετρία

Η τροπική άλγεβρα και η τροπική γεωμετρία αποτελούν γενικεύσεις των αντίστοιχων γραμμικών τους αναλόγων. Ο όρος *τροπικός* έχει αποδοθεί από Γάλλους μαθηματικούς προς τιμή του Βραζιλιάνου μαθηματικού Imre Simon, ο οποίος ήταν ενεργός ερευνητής στον τομέα. Απλοϊκά, η κύρια διαφορά της τροπικής άλγεβρας και της τροπικής γεωμετρίας από τα γραμμικά ανάλογα έγκειται στο γεγονός ότι το ζεύγος των πράξεων $(+, \times)$ αντικαθίσταται από το ζεύγος $(\min, +)$. Αυτό συνεπάγεται ότι αλλάζει η δομή όλων των πράξεων που εμπεριέχουν αυτές τις πράξεις, και ειδικά αλλάζει η δομή του πολλαπλασιασμού πινάκων.

2.1 Γραμμική Άλγεβρα

Προτού ξεκινήσουμε την ανάλυση της τροπικής άλγεβρας, κρίνεται σκόπιμο να αφιερώσουμε μια μικρή ενότητα σε μια σύντομη εισαγωγή στο γραμμικό της ανάλογο, την *γραμμική άλγεβρα* ([26], [24], [51]). Η γραμμική άλγεβρα αποτελεί ένα από τα σημαντικότερα επιστημονικά πεδία, καθώς αποτελεί βάση πάνω στην οποία έχουν ανθίσει πληθώρα άλλων ερευνητικών πεδίων, και αποτελεί, επίσης, τρόπο συνοπτικής γραφής, και λύσης, σύνθετων μαθηματικών εξισώσεων. Επιπλέον, η γραμμική άλγεβρα εισάγει την ιδέα της αναπαράστασης των εξισώσεων μέσω πινάκων και διανυσματικών χώρων, δίνοντας μια νέα διάσταση στην κατανόηση των προβλημάτων.

Η γραμμική άλγεβρα ασχολείται με συστήματα γραμμικών εξισώσεων της μορφής:

$$\mathbf{Ax} = \mathbf{b} \Leftrightarrow \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad (2.1)$$

Τα συστήματα αυτά δύναται να αναλυθούν γεωμετρικά χρησιμοποιώντας την γραμμική άλγεβρα. Πράγματι, εάν θεωρήσουμε κάθε στήλη του συστήματος ως ένα διάνυσμα, τότε αυτά δημιουργούν έναν διανυσματικό χώρο, του οποίου οι ιδιότητες μπορούν να μελετηθούν μέσω της γραμμικής άλγεβρας.

2.2 Τροπική Άλγεβρα

Η τροπική άλγεβρα αποτελεί ένα μαθηματικό σώμα παρεμφερές με την γραμμική άλγεβρα. Συγκεκριμένα, η τροπική άλγεβρα δρα πάνω σε αντικείμενα τα οποία ονομάζονται *lattices* ([23], [4], [16]). Αποτελεί έναν ημιδακτύλιο πάνω στο σύνολο αριθμών \mathbb{R}_{\min} . Το ζεύγος των θεμελιωδών πράξεων είναι το $(\min, +)$. Θα συμβολίζουμε το \min με το σύμβολο \wedge και τον \min -plus πολλαπλασιασμό πινάκων και διανυσμάτων με \boxplus . Στην συγκεκριμένη άλγεβρα

αλλάζει ο τρόπος με τον οποίο εκτελείται ο παραπάνω πολλαπλασιασμός. Συγκεκριμένα, ανάλογα με την γραμμική άλγεβρα, τα στοιχεία του γινομένου παίρνουν την μορφή:

$$(\mathbf{A} \boxplus \mathbf{B})_{ij} = \bigwedge_{k=1}^n A_{ik} + B_{kj} \quad (2.2)$$

Δηλαδή κάθε στοιχείο του γινομένου αποτελείται από το ελάχιστο των επιμέρους αθροισμάτων, κατ' αντιστοιχία της γραμμικής άλγεβρας όπου το κάθε στοιχείο αποτελείται από το άθροισμα των επιμέρους γινομένων (αφού εκεί το ζεύγος των θεμελιωδών πράξεων είναι το $(+, \times)$). Επιπλέον, σημαντική διαφοροποίηση της τροπικής από τη γραμμική άλγεβρα παρουσιάζεται μέσω της αρχής της επαλληλίας. Συγκεκριμένα, η τροπική άλγεβρα υπακούει την αρχή της \min επαλληλίας, ενώ η γραμμική άλγεβρα την αρχή της γραμμικής επαλληλίας. Γενικότερα, μπορούμε να συνοψίσουμε μερικές από τις κύριες διαφορές μεταξύ της τροπικής και της γραμμικής άλγεβρας στον Πίνακα 2.1.

Διαφορές γραμμικής και τροπικής άλγεβρας		
Στοιχείο	Γραμμική Άλγεβρα	Τροπική Άλγεβρα
Άθροισμα	+	\min
Πολλαπλασιασμός	\times	+
Ουδέτερο Πολλ/μου	1	0
Μηδενικό Πολλ/μου	0	∞

Πίνακας 2.1: Χαρακτηριστικές διαφορές θεμελιωδών στοιχείων μεταξύ της γραμμικής και της τροπικής άλγεβρας.

Υπάρχουν δύο πίνακες οι οποίοι έχουν πολύ μεγάλη σημασία στο πλαίσιο της τροπικής άλγεβρας. Έστω ένας πίνακας \mathbf{A} . Τότε, μπορούμε να ορίσουμε τους εξής δύο πίνακες:

- $\Gamma(\mathbf{A}) = \mathbf{A} \wedge \mathbf{A}^2 \wedge \dots \wedge \mathbf{A}^n \wedge \dots = \bigwedge_{n=1}^{\infty} \mathbf{A}^n$
- $\Delta(\mathbf{A}) = \mathbf{I} \wedge \mathbf{A} \wedge \mathbf{A}^2 \wedge \dots \wedge \mathbf{A}^n \wedge \dots = \bigwedge_{n=0}^{\infty} \mathbf{A}^n$

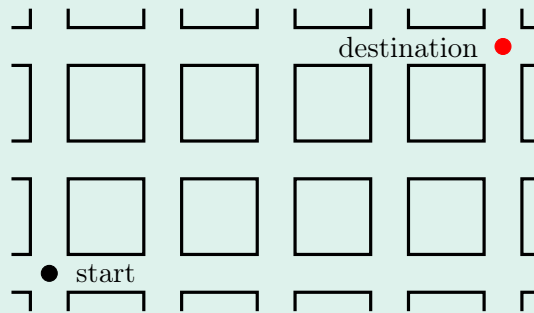
Αυτοί οι δύο πίνακες είναι ιδιαίτερα σημαντικοί στην τροπική άλγεβρα, καθώς παρέχουν λύσεις στα θεμελιώδη προβλήματα ιδιοτιμών. Συγκεκριμένα:

- Ο πίνακας $\Gamma(\mathbf{A})$ παρέχει λύσεις στο πρόβλημα $\mathbf{A} \boxplus \mathbf{x} = \lambda \boxplus \mathbf{x}$.
- Ο πίνακας $\Delta(\mathbf{A})$ παρέχει λύσεις στο πρόβλημα $\mathbf{A} \boxplus \mathbf{x} \geq \lambda \boxplus \mathbf{x}$.

Πριν ξεκινήσουμε την αυστηρή ανάλυση της τροπικής άλγεβρας, κρίνουμε σκόπιμο να παρουσιάσουμε μερικά ενδιαφέροντα παραδείγματα στα οποία προκύπτει οργανικά η ανάγκη χρήσης της τροπικής άλγεβρας. Σε πρώτη επαφή αυτή η μη-γραμμική άλγεβρα μπορεί να φαντάζει αυθαίρετη, αλλά προκύπτει φυσικά σε πληθώρα προβλημάτων. Τα παραδείγματα που θα παρουσιάσουμε στην ενότητα έχουν βασιστεί σε παραδείγματα από το σύγγραμμα του Butkovic ([10]).

Παράδειγμα 1. Πολλά σύγχρονα προβλήματα μπορούν να μοντελοποιηθούν χρησιμοποιώντας τη θεωρία γράφων. Ένα τέτοιο πρόβλημα με το οποίο αρκετοί άνθρωποι έρχονται σε επαφή σε καθημερινή βάση είναι η χρήση εφαρμογών χαρτών και συστημάτων πλοήγησης. Σε τέτοια προγράμματα οι διασταυρώσεις των δρόμων μοντελοποιούνται ως κόμβοι ενός γράφου και οι μεταξύ τους δρόμοι μοντελοποιούνται ως ακμές.

Σε αυτές τις ακμές ανατίθενται βάρη, τα οποία μπορεί να εξαρτώνται από την υπάρχουσα κίνηση, την ταχύτητα με την οποία μπορεί να διαβεί το όχημα τον συγκεκριμένο δρόμο, την συνολική απόσταση, κτλ. Το πρόβλημα που καλούνται να επιλύσουν τα συστήματα πλοήγησης είναι ο εντοπισμός της βέλτιστης (ανάλογα με τα κριτήρια του χρήστη) διαδρομής μεταξύ των κόμβων αφετηρίας και προορισμού. Σε αυτό το πλαίσιο προκύπτουν οργανικά οι τελεστές της τροπικής άλγεβρας, καθώς τα προγράμματα πλοήγησης επιθυμούν να υπολογίσουν το ελάχιστο άθροισμα βαρών, με βάση τα κριτήρια του χρήστη, το οποίο να συνδέει τον αρχικό με τον τελικό κόμβο.



Σχήμα 2.1: Παράδειγμα εύρεσης ελάχιστου μονοπατιού.

Τα προβλήματα ελαχίστων μονοπατιών έχουν επιπλέον ενδιαφέρον για την τροπική άλγεβρα, πέραν απλώς της ανάδυσης των τελεστών της. Όπως θα δούμε παρακάτω, η λύση του προβλήματος των ελαχίστων μονοπατιών αποτελεί ένα από τα κυριότερα προβλήματα με τα οποία καταπιάνεται η τροπική άλγεβρα, και η λύση τους προκύπτει αβίαστα μέσω των θεμελιωδών της πράξεων.

Παράδειγμα 2. Η \max -plus άλγεβρα αποτελεί ένα ισομορφισμό της τροπικής άλγεβρας. Οι δύο συνδέονται μέσω της απεικόνισης:

$$f : \mathbb{R}_{\min} \rightarrow \mathbb{R}_{\max}, f(\mathbf{x}) = -\mathbf{x}$$

Συνεπώς θα παρουσιάσουμε το παράδειγμα στην \max -plus άλγεβρα ώστε να διατηρείται η διαίσθηση. Το σημαντικό είναι ότι προκύπτουν οι συγκεκριμένοι μη-γραμμικοί τελεστές.

Έστω ότι ένα τρένο και ένα λεωφορείο φτάνουν στο αεροδρόμιο, από το οποίο αναχωρούν δύο πτήσεις. Το αεροδρόμιο, για να ελαχιστοποιήσει τον εκνευρισμό των ταξιδιωτών, αλλά και των αεροπορικών εταιριών, πρέπει να μελετήσει ποια θα πρέπει να είναι τα δρομολόγια των τρένων και των λεωφορείων, έτσι ώστε οι επιβάτες να προλάβουν να επιβιβαστούν στις πτήσεις. Ας συμβολίσουμε με λ_{g_1} , λ_{g_2} τους χρόνους που χρειάζονται οι επιβάτες για να μεταβούν από το σταθμό των λεωφορείων στις πύλες των δύο πτήσεων, και αντίστοιχα συμβολίζουμε ως τ_{g_1} , τ_{g_2} τους χρόνους που απαιτούνται για τη μετάβαση από τον σταθμό των τρένων. Τέλος, έστω ότι οι χρόνοι αναχώρησης των δύο πτήσεων είναι d_1 και d_2 αντίστοιχα. Θέλουμε να υπολογίσουμε τους χρόνους x_1 και x_2 που πρέπει να έχουν φτάσει, το πολύ, το λεωφορείο και το τρένο έτσι ώστε οι επιβάτες να μην χάσουν τις πτήσεις. Συνεπώς, τα μεγέθη πρέπει να ικανοποιούν τις παρακάτω

εξισώσεις:

$$\begin{aligned}\max(\lambda_{g_1} + x_1, \tau_{g_1} + x_2) &\leq d_1 \\ \max(\lambda_{g_2} + x_1, \tau_{g_2} + x_2) &\leq d_2\end{aligned}$$

Παρατηρούμε ότι θέτοντας:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \lambda_{g_1} & \tau_{g_1} \\ \lambda_{g_2} & \tau_{g_2} \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

τότε, οι εξισώσεις οι οποίες πρέπει να ικανοποιούν οι χρόνοι άφιξης του λεωφορείου και του τρένου δύναται να απλοποιηθούν ως εξής:

$$\mathbf{A} \boxplus \mathbf{x} \leq \mathbf{d}$$

Υπάρχουν δύο τύποι θεμελιώδης τύποι προβλημάτων οι οποίοι απαντώνται στην τροπική άλγεβρα, όπως υποδηλώνεται από τα παραπάνω παραδείγματα. Αυτά είναι τα προβλήματα ιδιοτιμών, και τα προβλήματα ελαχίστων αποστάσεων. Για να καταλάβουμε γιατί αυτά τα προβλήματα αποτελούν θεμελιώδη προβλήματα στην τροπική άλγεβρα, αλλά και να κάνουμε μια εισαγωγή στον τρόπο λύσης τους, θα διατυπώσουμε, και μετέπειτα θα αποδείξουμε, έναν ισχυρισμό αναφορικά με τον πίνακα $\Gamma(\mathbf{A})$.

Ισχυρισμός 2.1. Υποστηρίζουμε ότι ο πίνακας $\Gamma(\mathbf{A})$, εφόσον ο πίνακας \mathbf{A} αντιστοιχεί σε ένα δίκτυο χωρίς αρνητικούς κύκλους, είναι πεπερασμένος, και συγκεκριμένα ισούται με

$$\Gamma(\mathbf{A}) = \mathbf{A} \wedge \mathbf{A}^2 \wedge \dots \wedge \mathbf{A}^{n-1} \quad (2.3)$$

και εκφράζει τις ελάχιστες αποστάσεις μεταξύ όλων των κόμβων του δικτύου.

Για να αποδείξουμε τον παραπάνω ισχυρισμό χρειάζεται να ανατρέξουμε στην Θεωρητική Επιστήμη των Υπολογιστών, και συγκεκριμένα στον τομέα των Αλγορίθμων. Σύμφωνα με τη θεωρία των αλγορίθμων, εάν σε ένα δίκτυο που αποτελείται από, συνολικά, n κόμβους δεν υπάρχουν κύκλοι αρνητικού βάρους, τότε κάθε ελάχιστο μονοπάτι μεταξύ δύο οποιωνδήποτε κόμβων μπορεί να αποτελείται από το πολύ $n - 1$ ακμές. Ο λόγος γι' αυτό βασίζεται στην αρχή της βελτιστότητας, και αναλύεται στο ότι εάν έχουμε n (και, εν δυνάμει, παραπάνω) ακμές, τότε επισκεπτόμαστε τουλάχιστον $n + 1$ κόμβους (προσμετρώντας και τους κόμβους αφετηρίας και προορισμού). Όμως, αφού στο δίκτυο υπάρχουν συνολικά n διακριτοί κόμβοι, αυτό συνεπάγεται ότι θα έχουμε επισκεφθεί τουλάχιστον έναν κόμβο παραπάνω από μια φορά. Κανένα τέτοιο μονοπάτι, όμως, δεν γίνεται να είναι βέλτιστο. Έστω ότι το βέλτιστο μονοπάτι μεταξύ δύο κόμβων αποτελούνταν από ακριβώς n ακμές, και έστω v ένας κόμβος που επισκεπτόμαστε τουλάχιστον 2 φορές. Έστω l το μονοπάτι μετά την τελευταία επίσκεψη στον κόμβο v το οποίο καταλήγει, εν τέλει, στον κόμβο προορισμού. Συνολικά, αυτή η διαδρομή δεν είναι δυνατόν να είναι βέλτιστη, δεδομένου ότι το δίκτυο είναι στατικό και δεν εξελίσσεται στον χρόνο, καθώς έχουμε τη δυνατότητα να ακολουθήσουμε το μονοπάτι l που καταλήγει στον κόμβο προορισμού την πρώτη φορά που επισκεπτόμαστε τον κόμβο v , γλιτώνοντας έτσι την κυκλική διαδρομή μεταξύ των δύο επισκέψεων του κόμβου v . Αυτό συνεπάγεται ότι κανένα μονοπάτι το οποίο αποτελείται από n , και παραπάνω, ακμές δεν δύναται να είναι βέλτιστο.

Έχοντας παρουσιάσει μια σύντομη απόδειξη για το μήκος μιας ελάχιστης απόστασης, θα εξηγήσουμε πώς ο Ισχυρισμός 2.1 υπολογίζει τα ελάχιστα μονοπάτια. Η n -ιοστή δύναμη του πίνακα \mathbf{A} , πρακτικά, υπολογίζει τις αποστάσεις μεταξύ των κόμβων του δικτύου για μονοπάτια

που αποτελούνται από ν ακμές. Συνεπώς, η απόδειξη του Ισχυρισμού 2.1 έπεται από το γεγονός ότι ο πίνακας $\Gamma(\mathbf{A})$ αποτελείται, για κάθε ζεύγος κόμβων, από την ελάχιστη τιμή μεταξύ των μονοπατιών για όλα τα δυνατά μήκη μονοπατιού. Ο υπολογισμός που επιτελεί ο πίνακας ταυτίζεται με έναν θεμελιώδη αλγόριθμο της Επιστήμης των Υπολογιστών, τον αλγόριθμο Floyd-Warshall.

Συνεπώς, καταλαβαίνουμε ότι τα προβλήματα ελάχιστων αποστάσεων δύναται να μοντελοποιηθούν οργανικά στην τροπική άλγεβρα. Επιπλέον, ένα από τα πιο θεμελιώδη μεγέθη του πεδίου είναι αφιερωμένο στην άμεση, και εύκολη, επίλυση τέτοιων προβλημάτων. Προχωρώντας στον δεύτερο τύπο προβλημάτων, έχουμε τα προβλήματα γενικευμένων ιδιοτιμών. Αυτά τα προβλήματα, όπως αναφέραμε και προηγουμένως, είναι της μορφής:

$$\mathbf{A} \boxplus \mathbf{x} \geq \lambda \boxplus \mathbf{x} \quad (2.4)$$

Αξίζει, όμως, να αναλογιστούμε τι ακριβώς κρύβουν οι εξισώσεις των ιδιοτιμών και τι ιδιαίτερο έχουν τα προβλήματα τα οποία λύνουν. Αρχικά, οι εξισώσεις της μορφής της Εξίσωσης (2.4) εξασφαλίζουν ότι το ελάχιστο των στοιχείων του γινομένου των ποσοτήτων είναι μεγαλύτερο από κάποια τιμή. Όμως, από τη στιγμή που το ελάχιστο στοιχείο είναι μεγαλύτερο από μια τιμή, αυτό συνεπάγεται ότι όλα τα στοιχεία είναι μεγαλύτερα από την τιμή αυτή. Κάτι τέτοιο μπορεί να αποβεί χρήσιμο σε πληθώρα προβλημάτων. Στον σχεδιασμό συστημάτων, για παράδειγμα, μπορούμε να σχεδιάσουμε κατάλληλα τη μοντελοποίηση των πινάκων μαζί με τις παραμέτρους του συστήματος, και μαζί με τις επιθυμητές συμπεριφορές, και έπειτα, επιλύοντας το πρόβλημα των ιδιοτιμών, να αποφανθούμε εάν είναι δυνατό να σχεδιάσουμε ένα σύστημα το οποίο να έχει ακριβώς την επιθυμητή συμπεριφορά (ή, στην περίπτωση του πίνακα $\Delta(\mathbf{A})$, να εξασφαλίσουμε ότι το σύστημα θα έχει το πολύ κάποια συγκεκριμένη συμπεριφορά).

Διαφορετικά, σε ένα διαισθητικό επίπεδο, αξίζει να αναρωτηθούμε τι ακριβώς σημαίνουν οι εξισώσεις ιδιοτιμών. Προς αυτόν τον σκοπό θα επιστρατεύσουμε, και πάλι, τη γραμμική άλγεβρα, και συγκεκριμένα τη θεωρία των διανυσματικών χώρων. Σε αυτό το πλαίσιο, οι πίνακες δεν είναι παρά τελεστές οι οποίοι αλλοιώνουν, με κάποιο τρόπο, τα διανύσματα εφαρμογής, και μπορούμε, καταχρηστικά, να τους σκεφτόμαστε σαν συναρτήσεις που επιδρούν πάνω σε εισόδους. Ένα χαρακτηριστικό παράδειγμα έρχεται τέτοιας θεώρησης προέρχεται από το πεδίο της ρομποτικής, όπου οι ορθοκανονικοί πίνακες διαστάσεων 3×3 , οι οποίοι ορίζουν την ομάδα των ειδικών ορθογωνικών πινάκων $SO(3)$ έχουν ως αποτέλεσμα την περιστροφή ενός διανύσματος στο χώρο, σύμφωνα με κάποιοι άξονα. Υπό αυτό το πρίσμα των διανυσματικών χώρων, οι εξισώσεις ιδιοτιμών έχουν μια πολύ ευχάριστη ιδιότητα: ποσοτικοποιούν το κατά πόσο ο τελεστής αλλοιώνει την είσοδο, και, στην περίπτωση της ακριβής λύσης, προσδιορίζουν επακριβώς τη λειτουργία του τελεστή.

Μια ενδιαφέρουσα σύνδεση που μπορεί να κάνει κανείς είναι η αναλογία των προβλημάτων ιδιοτιμών με τη συνέχεια συναρτήσεων κατά Lipschitz. Ακριβώς όπως η εν λόγω συνέχεια για μια συνάρτηση εκφράζει ότι η επίδραση της συνάρτησης στις εισόδους μπορεί να έχει μέχρι μια πεπερασμένη επιρροή, έτσι ακριβώς και τα προβλήματα ιδιοτιμών εκφράζουν ότι υπάρχουν εισόδοι για τις οποίες η λειτουργία των τελεστών είναι πεπερασμένη, και ελεγχόμενη. Αν και οι έννοιες είναι παρόμοιες, υπάρχει μια σημαντική, και καταλυτική διαφορά: η συνέχεια κατά Lipschitz είναι, συνήθως, καθολική και ισχύει για όλες τις πιθανές εισόδους. Αντίθετα, τα προβλήματα ιδιοτιμών είναι κατά κανόνα εφαρμόσιμα σε ένα μικρό υποσύνολο του χώρου εισόδου. Όμως, επιτρέπουν στον χρήστη να καθορίσει την συμπεριφορά του συστήματος σε αυτές τις περιπτώσεις. Έτσι, εάν καταφέρει να περιορίσει τον χώρο των δυνατών εισόδων σε ένα πεπερασμένο (με μικρές αποκλίσεις) σύνολο ιδιοδιανυσμάτων, τότε η συμπεριφορά του συστήματος είναι πλήρως ελεγχόμενη.

2.3 Τροπική Γεωμετρία

Η τροπική γεωμετρία ([39]) αποτελείται από την εφαρμογή των ιδεών των πολυέδρων και των πολυτόπων ([58], [21]) της αλγεβρικής γεωμετρίας ([28], [25], [27], [49]) στον φορμαλισμό της τροπικής άλγεβρας. Τα κύρια αντικείμενα της τροπικής γεωμετρίας είναι οι ημιχώροι, τα πολυέδρα, και τα πολύτοπα.

Ας εξετάσουμε λίγο τους παραπάνω όρους:

- ένας ημιχώρος ορίζεται από την εφαρμογή μιας ανισότητας σε έναν αυθαίρετο d -διάστατο χώρο. Έστω η ανισότητα:

$$x \leq 5,$$

ορισμένη στο γνωστό μας επίπεδο (δισδιάστατος χώρος). Αυτή η ανισότητα ορίζει έναν ημιχώρο ο οποίος περιέχει όλα τα σημεία τα οποία έχουν την x συντεταγμένη μικρότερη του 5, ανεξάρτητα με την τιμή της y συντεταγμένης τους.

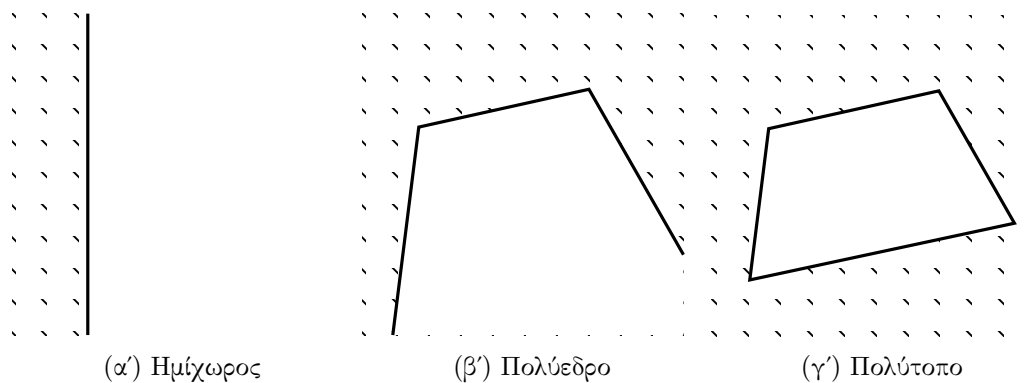
- ένα πολυέδρο προκύπτει από τον συνδυασμό ενός πεπερασμένου αριθμού ημιχώρων. Επεκτείνοντας το προηγούμενο παράδειγμα, ας προσθέσουμε επιπλέον τον περιορισμό ότι η y συντεταγμένη πρέπει να είναι μικρότερη του 6. Επίσημα:

$$\begin{cases} x \leq 5 \\ y \leq 6 \end{cases}$$

- τέλος, ένα πολύτοπο είναι απλώς ένα πολυέδρο το οποίο είναι φραγμένο. Δηλαδή οι εξισώσεις οι οποίες το ορίζουν δημιουργούν ένα κλειστό σχήμα το οποίο δεν εκτείνεται στο άπειρο. Θεωρώντας πάλι το παράδειγμά μας, εάν εμπλουτίσουμε τους περιορισμούς μας με δύο ακόμα εξισώσεις μπορούμε να ορίσουμε ένα πολύτοπο. Συγκεκριμένα:

$$\begin{cases} 0 \leq x \leq 5 \\ -2 \leq y \leq 6 \end{cases}$$

Γραφικές αναπαραστάσεις των παραπάνω όρων δίνονται στο Σχήμα 2.2.



Σχήμα 2.2: Αναπαράσταση ενός Ευκλείδειου ημιχώρου, ενός πολυέδρου, και ενός πολυτόπου.

Για να φέρουμε τον αναγνώστη σε μεγαλύτερη εξοικείωση με την τροπική γεωμετρία και να τον βοηθήσουμε να κατανοήσει σε μεγαλύτερο βάθος τις κύριες διαφορές της από την αναλυτική γεωμετρία θα εξετάσουμε μερικά παραδείγματα που υπογραμμίζουν την διαφορετική συμπεριφορά τους. Αρχικά, θα εξετάσουμε την μορφή που αποκτά η εξίσωση της ευθείας στο

πλαίσιο της τροπικής άλγεβρας. Ως γνωστόν, μια ευθεία αποτελείται από το σύνολο των σημείων τα οποία ικανοποιούν την εξίσωσή της. Η ευθεία μπορεί να καθοριστεί πλήρως από δύο παραμέτρους, την κλίση της, α , και την τομή της, β :

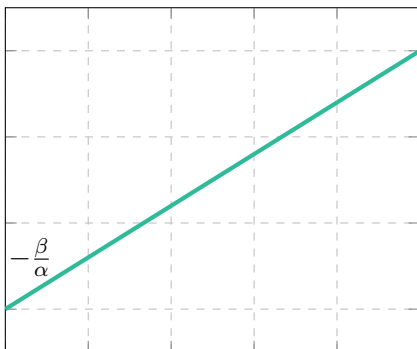
$$y = \alpha \cdot x + \beta \quad (2.5)$$

Η παράμετρος της κλίσης καθορίζει το πόσο απότομα αυξάνονται οι τιμές της ευθείας, και η παράμετρος της τομής καθορίζει το σημείο τομής της ευθείας με τον άξονα των τεταγμένων. Συνεπώς, το σύνολο των σημείων του διδιάστατου χώρου των οποίων οι τεταγμένες και οι τεταγμένες ικανοποιούν την Εξίσωση (2.5) αποτελούν την ίδια την ευθεία.

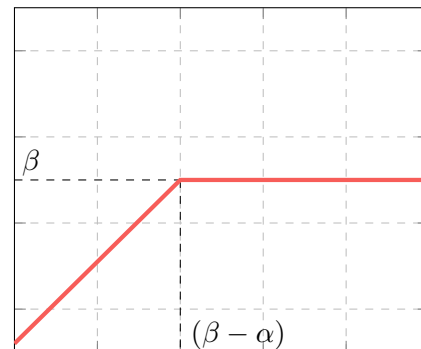
Όταν μεταφερόμαστε στο πεδίο της τροπικής άλγεβρας δεν μεταβάλλεται ο ορισμός της ευθείας. Όμως, καθώς η αναλυτική γεωμετρία βασίζεται στους τελεστές της γραμμικής άλγεβρας, ενώ η τροπική γεωμετρία στους τελεστές της τροπικής άλγεβρας, η μορφή της ευθείας στα δύο πεδία είναι παντελώς διαφορετική. Συγκεκριμένα, εκφράζοντας την Εξίσωση (2.5) στο πλαίσιο της τροπικής άλγεβρας έχουμε:

$$y = (\alpha + x) \wedge \beta = \min(\alpha + x, \beta) \quad (2.6)$$

Ο μη-γραμμικός τελεστής \min περιμένουμε ότι θα εισάγει κάποιο 'σπάσιμο' στην γραφική παράσταση της Εξίσωσης (2.6). Επιπλέον, καθώς τα ορίσματα του τελεστή είναι 2 στον αριθμό και είναι γραμμικά, περιμένουμε ότι η γραφική παράσταση θα είναι τμηματικά γραμμική, με δύο χωριστά γραμμικά τμήματα. Το Σχήμα 2.3 επιβεβαιώνει τις παραπάνω παρατηρήσεις και παρουσιάζει συγκεντρωτικά τις γραφικές παραστάσεις των Εξισώσεων (2.5) και (2.6).



(α') $y = \alpha \cdot x + \beta$



(β') $y = \min(\alpha + x, \beta)$

Σχήμα 2.3: Αναπαράσταση της γενικής εξίσωσης της ευθείας στην αλγεβρική και τροπική γεωμετρία.

Ως ένα δεύτερο παράδειγμα για την εξοικείωση του αναγνώστη, θα επιστρατεύσουμε την εξίσωση της παραβολής. Αντίστοιχα με την ευθεία, μια παραβολή αποτελείται από το σύνολο των σημείων που ικανοποιούν την εξίσωσή της. Όντας, όμως, εξίσωση μεγαλύτερης τάξης απαιτεί και περισσότερες παραμέτρους για την ακριβή περιγραφή της. Συγκεκριμένα, η εξίσωση της παραβολής έχει την εξής μορφή:

$$y = \alpha \cdot x^2 + \beta \cdot x + \gamma \quad (2.7)$$

Οι παράμετροι της Εξίσωσης (2.7) επηρεάζουν την μορφή της γραφικής παράστασης. Συγκεκριμένα, σε γενικές γραμμές:

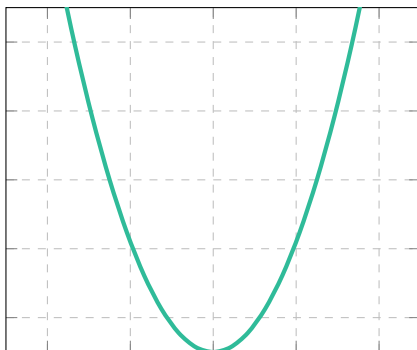
- η παράμετρος α καθορίζει το πόσο απότομα αυξάνονται οι τιμές της παραβολής. Μεγαλύτερες τιμές της παραμέτρου οδηγούν σε πιο ‘στενή’ γραφική παράσταση (και συνεπώς σε μεγαλύτερο ρυθμό αύξησης).
- η παράμετρος β καθορίζει τα σημεία όπου η παράσταση τέμνει των άξονα των τετμημένων.
- η παράμετρος γ καθορίζει την μετατόπιση στον άξονα των τεταγμένων.

Η μορφή της παραβολής δεν είναι τόσο απλή όσο υπονοείται από την επεξήγηση των παραμέτρων της. Οι παράμετροι δεν επιδρούν ανεξάρτητα στην γραφική παράσταση, αλλά σε συνδυασμό, και έτσι οι αλληλεπιδράσεις τους καθορίζουν την τελική μορφή της παραβολής.

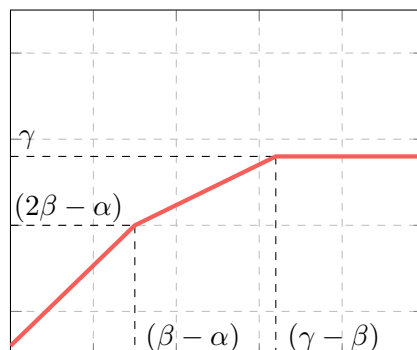
Στο πεδίο της τροπικής άλγεβρας η εξίσωση της παραβολής έχει δραστικότερες διαφορές από την εξίσωση της παραβολής στην γραμμική άλγεβρα. Στην γραμμική άλγεβρα, ο δευτεροβάθμιος όρος $\alpha \cdot x^2$ εισάγει μια μη-γραμμικότητα στην γραφική παράσταση. Στην τροπική άλγεβρα, οι μη-γραμμικότητες είναι έμφυτο στοιχείο, όμως η έκφασή τους είναι διαφορετική: οι μη-γραμμικότητες εμφανίζονται ως ‘σπασίματα’ σε μια, κατά τα άλλα, γραμμική γραφική παράσταση. Αυτό αποτελεί σημείο μεγάλου ενδιαφέροντος, και μια από τις κυριότερες διαφορές μεταξύ της τροπικής και της γραμμικής άλγεβρας: αντίθετα με την γραμμική άλγεβρα, όπου αυξάνοντας την τάξη μιας εξίσωσης εισάγονται μη-γραμμικότητες νέου τύπου, στην τροπική άλγεβρα η αύξηση της τάξης μιας εξίσωσης εισάγει απλώς ένα επιπλέον γραμμικό τμήμα. Αυτή η ιδιότητα της τροπικής γεωμετρίας αποτελεί και έναν από τους κυριότερους λόγους που έχει γίνει τόσο ελκυστικό αντικείμενο μελέτης. Η μορφή αυθαίρετα μεγάλων τροπικών πολυωνύμων είναι γνωστή, και αποτελείται πάντα από έναν πεπερασμένο αριθμό γραμμικών τμημάτων. Αυτό το γεγονός είναι εξαιρετικά χρήσιμο στην ανάλυση της πολυπλοκότητας των αλγορίθμων, και περιορίζει σημαντικά τον χώρο πιθανών λύσεων ενός προβλήματος, καθώς απαιτείται μονάχα ο έλεγχος ενός πεπερασμένου αριθμού σημείων. Συγκεκριμένα, σε τέτοια προβλήματα οι πιθανές λύσεις δύνανται να βρίσκονται μονάχα σε σημεία τομής γραμμικών περιοχών, αποτέλεσμα το οποίο προκύπτει από την περιοχή του γραμμικού προγραμματισμού ([12], [34], [38]). Έχοντας κάνει αυτή την εκτενή εισαγωγή, παραθέτουμε την έκφραση της Εξίσωσης (2.7) στο πλαίσιο της τροπικής άλγεβρας:

$$y = (\alpha + 2 \cdot x) \wedge (\beta + x) \wedge \gamma = \min(\alpha + 2 \cdot x, \beta + x, \gamma) \quad (2.8)$$

Στο Σχήμα 2.4 παρουσιάζουμε μια οπτική σύγκριση των παραβολών που ορίσαμε στις Εξισώσεις (2.7) και (2.8). Στην γραμμική αναπαράσταση της παραβολής δεν έχουμε επισημειώσει συγκεκριμένες τιμές, καθώς όπως αναφέραμε και στην ανάλυση της Εξίσωσης (2.7), η συγκεκριμένη μορφή εξαρτάται από την εκάστοτε αλληλεπίδραση των παραμέτρων. Το σχήμα υπογραμμίζει τις διαφορές που αναφέραμε στην παραπάνω ανάλυση, και εστιάζουμε την προσοχή του αναγνώστη στην δεύτερη παράσταση του Σχήματος 2.4 όπου είναι εμφανείς οι τρεις διακριτές γραμμικές περιοχές.



$$(\alpha') y = \alpha \cdot x^2 + \beta \cdot x + \gamma$$



$$(\beta') y = \min(\alpha + 2 \cdot x, \beta + x, \gamma)$$

Σχήμα 2.4: Αναπαράσταση της εξίσωσης παραβολής στην αλγεβρική και τροπική γεωμετρία.

Κεφάλαιο 3

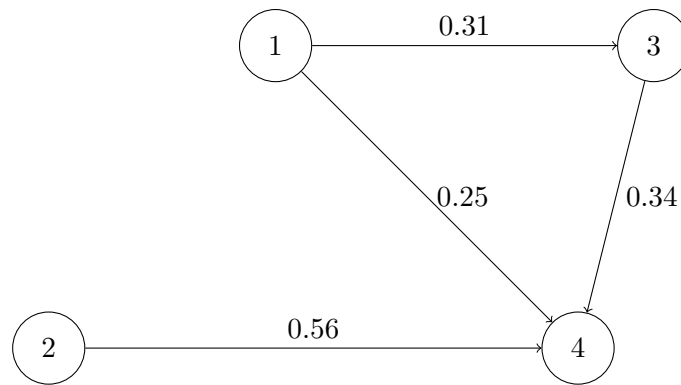
Γραφικά Μοντέλα

Στο Κεφάλαιο 2, αλλά και στην εισαγωγή, κάναμε αναφορές σε γράφους και γραφικά μοντέλα, χωρίς όμως να ορίσουμε μαθηματικά αυτά τα αντικείμενα, και χωρίς να κάνουμε νύξη στα διαφορετικά ήδη γράφων τα οποία απαντώνται στην Επιστήμη των Υπολογιστών. Θα αφιερώσουμε αυτό το κεφάλαιο σε μια σύντομη εξέταση του κόσμου των γραφικών μοντέλων, έτσι ώστε να έχουμε στη διάθεσή μας τα απαραίτητα μαθηματικά εργαλεία για μια ολοκληρωμένη μελέτη των αλγορίθμων που δρουν πάνω σε αυτά τα μοντέλα στα επόμενα κεφάλαια. Τα γραφικά μοντέλα αποτελούν δομές δεδομένων που επιτρέπουν την περιγραφή πληθώρα προβλημάτων στην Επιστήμη των Υπολογιστών. Επιπλέον, λόγω της έμφυτης δομής τους, επιδέχονται γραφική αναπαράσταση η οποία επιτρέπει την οπτική αναπαράσταση της πληροφορίας. Κάτι τέτοιο καθιστά την κατανόηση των εξεταζόμενων προβλημάτων σημαντικά απλούστερη, και επιτρέπει την εποπτική εξέταση του χώρου λύσεων. Συνοπτικά, τα διαφόρων λογής γραφικά μοντέλα αποτελούν δομές ικανές να αναπαραστήσουν, συνοπτικά και συμπυκνωμένα, την πληροφορία των προβλημάτων που μοντελοποιούν.

3.1 Γράφοι

Η πιο απλή και πιο πολυχρησιμοποιημένη μορφή γραφικών μοντέλων που απαντάται στην Επιστήμη των Υπολογιστών είναι οι *γράφοι* και αποτελούν το αντικείμενο μελέτης της Θεωρίας Γράφων ([22], [6], [56]). Απλουστευμένα, οι γράφοι αποτελούνται από ένα σύνολο κόμβων οι οποίοι συνδέονται μεταξύ τους με ακμές. Συνεπώς, οποιοδήποτε πρόβλημα εμπεριέχει στον πυρήνα του αντικείμενα τα οποία αλληλεπιδρούν μεταξύ τους μπορεί, σε κάποιο βαθμό, να μοντελοποιηθεί με τη Θεωρία Γράφων. Συγκεκριμένα, μέσω γράφων δύναται να μοντελοποιηθούν ένας συνταρακτικός αριθμός προβλημάτων, ειδικά στο πεδίο της Θεωρητικής Επιστήμης των Υπολογιστών. Ο λόγος γι' αυτό είναι ότι αν και οι γράφοι αποτελούν τη λιγότερο εκφραστική δομή συγκριτικά με τις υπόλοιπες τις οποίες θα εξετάσουμε, είναι εξαιρετικά ευέλικτοι. Αρκετές από τις δομές που θα δούμε παρακάτω αποτελούν επεκτάσεις των γράφων, όπου οι δομές έχουν εμπλουτιστεί με νέες δυνατότητες (συγκεκριμένα με σύμβολα εισόδου/εξόδου, καταστάσεις, κ.ά.). Στο Σχήμα 3.1 παρουσιάζεται ένας γράφος.

Οι γράφοι αποτελούν αρκετά λιτές δομές ως προς την περιγραφή τους. Συγκεκριμένα, για τον πλήρη μαθηματικό ορισμό ενός γράφου χρειάζεται να προσδιορίσουμε μονάχα δύο σύνολα: το σύνολο V των κόμβων του γράφου, και το σύνολο E των ακμών που συνδέουν τους κόμβους μεταξύ τους. Οι ακμές αναπαριστώνται με ζεύγη τις μορφής (u, v) , όπου $u, v \in V$. Έχοντας προσδιορίσει αυτά τα δύο σύνολα, ο γράφος είναι πλήρως ορισμένος. Η πιο συχνή, ίσως, αναπαράσταση των γράφων γίνεται μέσω του πίνακα πρόσπτωσης του γράφου, \mathbf{A} . Ο πίνακας πρόσπτωσης εμπεριέχει όλη την πληροφορία που χρειάζεται για τον προσδιορισμό του γράφου, καθώς μπορούν να προσδιοριστούν τόσο το σύνολο V των κόμβων (αντιστοιχώντας



Σχήμα 3.1: Ένας τυπικός γράφος.

τις γραμμές ή τις στήλες του πίνακα πρόσπτωσης στους κόμβους του συνόλου V), όσο και το σύνολο E των ακμών. Για παράδειγμα, ο πίνακας πρόσπτωσης για τον γράφο του Σχήματος 3.1 είναι:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0.31 & 0.25 \\ 0 & 0 & 0.56 & 0 \\ 0.31 & 0.56 & 0 & 0.34 \\ 0.25 & 0 & 0.34 & 0 \end{bmatrix}$$

Ο παραπάνω πίνακας χρίζει διευκρίνισης. Πράγματι, τόσο γράφος του Σχήματος 3.1 όσο και ο παραπάνω πίνακας πρόσπτωσης υποθέτουν έναν *μη κατευθυνόμενο* γράφο.

Ορισμός 3.1. Ένας γράφος θα ονομάζεται *κατευθυνόμενος* εάν τα ζεύγη των ακμών του είναι διατεταγμένα. Διαφορετικά, θα ονομάζεται *μη κατευθυνόμενος*.

Πρακτικά, στους κατευθυνόμενους γράφους, που θίγει ο Ορισμός 3.1, οι ακμές δεν είναι αμφίδρομες. Αυτό συνεπάγεται ότι η ύπαρξη της ακμής (u, v) σε έναν κατευθυνόμενο γράφο εκφράζει ότι δύναται η μεταφορά από τον κόμβο u στον v , αλλά όχι αντίστροφα (όπως, για παράδειγμα, συμβαίνει με τις μονοδρομήσεις). Η έννοια της κατεύθυνσης στους γράφους χρίζει αναφοράς καθώς αποτελεί δομικό λίθο των επόμενων δομών που θα αναλύσουμε.

Όπως αναφέραμε οι γράφοι χρησιμοποιούνται σε πληθώρα πεδίων εφαρμογών, τόσο σε προβλήματα της Επιστήμης των Υπολογιστών, όσο και σε προβλήματα αμιγώς των Μαθηματικών, της Βιολογίας, και των Κοινωνικών Επιστημών. Μερικά χαρακτηριστικά παραδείγματα τα οποία αντιπροσωπεύουν κατηγορίες προβλημάτων αποτελούν:

- Σε ένα οδικό δίκτυο, τα σταυροδρόμια μοντελοποιούνται ως οι κόμβοι του συστήματος και οι δρόμοι ως οι ακμές. Διαφορετικά, εάν εξετάζεται το δίκτυο σε μεγαλύτερη κλίμακα, είναι συχνό να αναπαριστώνται οι πόλεις ως κόμβοι, και οι αυτοκινητόδρομοι μεταξύ τους ως ακμές. Μέσω αυτής της μοντελοποίησης δύναται να επιλυθούν πολλά προβλήματα αναφορικά με τα οδικά δίκτυα. Χαρακτηριστικό παράδειγμα αποτελεί ο υπολογισμός συντομότερων διαδρομών πάνω στο χάρτη και η μετέπειτα κατάστρωση πιθανών διαδρομών.
- Σε ένα κοινωνικό δίκτυο οι χρήστες μπορούν να μοντελοποιηθούν ως κόμβοι, και οι ακμές μεταξύ τους εκφράζουν εάν οι χρήστες είναι συνδεδεμένοι μεταξύ τους. Σημειωτέον ότι ανάλογα το κοινωνικό δίκτυο, η παραπάνω σχέση μπορεί να είναι συμμετρική, και κατά συνέπεια και ο πίνακας πρόσπτωσης να είναι συμμετρικός, όπως για παράδειγμα στο

Facebook, αλλά μπορεί να είναι και κατευθυνόμενη, όπως για παράδειγμα στο Twitter. Μέσω αυτής της μοντελοποίησης δύναται να υπολογιστούν πολλά στατιστικά αναφορικά με το κοινωνικό δίκτυο, όπως για παράδειγμα να εκτελεστούν πειράματα διάδοσης πληροφορίας, ή υπολογισμός κλικών.

- Σε ένα δίκτυο όπου εξετάζεται η διάδοση μιας ασθένειας. Σε αυτό το δίκτυο οι άνθρωποι μοντελοποιούνται ως κόμβοι, και οι αλληλεπιδράσεις μεταξύ των ανθρώπων μοντελοποιούνται ως ακμές (με κατάλληλα βάρη, έτσι ώστε να συνυπολογίζεται και η μεταδοτικότητα της ασθένειας). Μέσω της μοντελοποίησης μπορεί να υπολογιστεί η βέλτιστη στρατηγική εμβολιασμού έτσι ώστε να εμποδιστεί η εξάπλωση της ασθένειας, ελαχιστοποιώντας παράλληλα το κόστος.

Έχοντας ολοκληρώσει τη σύντομη εισαγωγή στη Θεωρία Γράφων, καθώς και έχοντας παρουσιάσει πλήθος παραδειγμάτων όπου η Θεωρία Γράφων είναι άμεσα εφαρμόσιμη, θα αναφέρουμε μερικά από τα προβλήματα τα οποία απασχολούν αμιγώς το πεδίο. Ο λόγος γι' αυτό είναι, αφ' ενός, για την πληρέστερη παρουσίαση του πεδίου. Αφ' εταίρου, πολλές από τις πιθανές εφαρμογές που αναφέραμε παραπάνω (καθώς και κάποια από τα προβλήματα των επερχόμενων χωρίων), δύναται να αναχθούν σε κάποιο από αυτά. Συγκεκριμένα, τα μερικά από τα κύρια προβλήματα με τα οποία καταπιάνεται η Θεωρία Γράφων είναι:

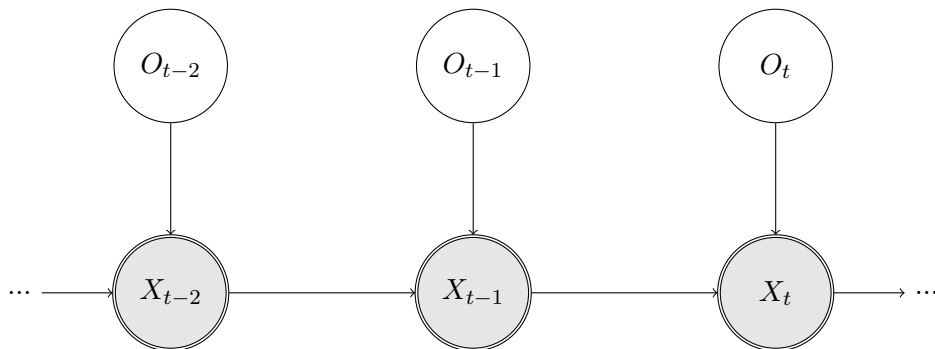
- *Χρωματισμός Γράφων.* Σε αυτό το πρόβλημα καλούμαστε να χρωματίσουμε τους κόμβους του γράφου με χρώματα έτσι ώστε να ισχύει κάποια ιδιότητα. Μπορεί, για παράδειγμα, να χρησιμοποιηθεί σε περίπτωση που προσπαθούμε να σχεδιάσουμε ένα γαμήλιο τραπέζι: πρέπει να υπολογιστούν τα χρώματα (που αντιστοιχούν στα τραπέζια) που πρέπει να χρωματίσουμε τους κόμβους (που αντιστοιχούν στους καλεσμένους), έτσι ώστε κόμβοι που συνδέονται με θετικά βάρη (υποδηλώνοντας φιλική σχέση) να έχουν το ίδιο χρώμα, ενώ παράλληλα να αποφεύγουμε να χρωματίζουμε με ίδιο χρώμα κόμβους που συνδέονται με αρνητικά βάρη (υποδηλώνοντας εχθρική σχέση).
- *Ισομορφισμός Υπογραφών.* Σε αυτό το πρόβλημα καλούμαστε εντοπίσουμε εάν απαντάται σε έναν γενικότερο γράφο κάποιο ειδικότερο πρότυπο. Μπορεί να χρησιμοποιηθεί εάν θέλουμε, για παράδειγμα, να εντοπίσουμε συγκεκριμένες δομές ενδιαφέροντος στο δίκτυο. Χαρακτηριστικό παράδειγμα απαντάται στον τομέα της Χημείας, όπου υπάρχει προσπάθεια για εντοπισμό κοινών χαρακτηριστικών σε χημικά στοιχεία.
- *Εντοπισμός Ελάχιστων Μονοπατιών.* Σε αυτό το πρόβλημα καλούμαστε να υπολογίσουμε τα συντομότερα μονοπάτια μεταξύ των ζευγών κόμβων του γράφου. Αποτελεί ένα από τα πιο δημοφιλή προβλήματα της Επιστήμης των Υπολογιστών στο σύνολό της. Χαρακτηριστικό παράδειγμα εφαρμογής αποτελούν τα συστήματα πλοήγησης, όπου καλείται να υπολογιστεί η ελάχιστη, ανάλογα με τα κριτήρια του χρήστη, διαδρομή μεταξύ δύο σημείων.
- *Υπολογισμός Ροής.* Σε αυτό το πρόβλημα τα βάρη των ακμών αντιμετωπίζονται ως χωρητικότητες. Καλούμαστε να υπολογίσουμε τη μέγιστη ροή που μπορεί να διέλθει μεταξύ δύο κόμβων του δικτύου. Αποτελεί, επίσης, ένα ιδιαίτερα δημοφιλές πρόβλημα της Επιστήμης των Υπολογιστών. Χαρακτηριστικό παράδειγμα απαντάται στον σχεδιασμό τηλεπικοινωνιακών δικτύων, όπου οι σχεδιαστές του συστήματος υπολογίζουν πιο είναι το μέγιστο φορτίο που μπορεί να υποστηρίξει το δίκτυο, και μετέπειτα αποφασίζουν εάν αυτό είναι ικανό για να ικανοποιήσει τις προδιαγραφές τους.

3.2 Hidden Markov Models

Τα Hidden Markov Models (HMMs) ([11], [47]) αποτελούν μια πολύ ιδιαίτερη κλάση γραφικών μοντέλων. Παρόμοια με τους γράφους, η αναπαράστασή τους αποτελείται από κόμβους οι οποίοι συνδέονται με ακμές. Εκεί, όμως, παύει η παρατήρηση ομοιοτήτων μεταξύ των δομών. Αντίθετα με τους γράφους, τα HMM επιδέχονται εισόδους, με τη μορφή παρατηρήσεων. Συγκεκριμένα, κάθε κόμβος έχει ξεχωριστή πιθανότητα να παρατηρήσει τα διαφορετικά σύμβολα εισόδου (παρατηρήσεις). Επιπλέον, οι κόμβοι επιδέχονται διαφορετικής επεξήγησης στο πλαίσιο των HMM συγκριτικά με τους γράφους. Ειδικότερα, στα Μαρκοβιανά μοντέλα οι κόμβοι χωρίζονται σε δύο βασικές κατηγορίες:

- **κόμβοι καταστάσεων:** αποτελούν κόμβους που εκφράζουν την κατάσταση του δικτύου. Από αυτούς τους κόμβους παίρνουν και την ονομασία των κρυφών μοντέλων. Αυτό προκύπτει από το γεγονός ότι υποθέτουμε ότι οι κόμβοι κατάστασης δεν είναι άμεσα παρατηρήσιμοι, και συνεπώς δεν είναι γνωστή η πραγματική κατάσταση του συστήματος.
- **κόμβοι παρατηρήσεων:** αποτελούν κόμβους που μοντελοποιούν τις παρατηρήσεις του συστήματος, οι οποίες αποτελούν τις εισόδους των HMMs. Μέσω των παρατηρήσεων γίνεται προσπάθεια, χρησιμοποιώντας τους αλγορίθμους των HMMs, να εκτιμηθεί η πραγματική κατάσταση του δικτύου.

Καταλαβαίνουμε, λοιπόν, ότι τα Μαρκοβιανά μοντέλα παρουσιάζουν έμφυτες διαφορές από τους γράφους. Συγκεκριμένα, στους γράφους οι κόμβοι εκφράζουν την ύπαρξη κάποιου στοιχείου (δηλαδή κάποιου χρήστη, κάποιας πόλης) και οι ακμές εκφράζουν τον τρόπο με τον οποίο συνδέονται (ή αλληλεπιδρούν) τα στοιχεία. Αντίθετα, στα HMMs οι κόμβοι είτε εκφράζουν διακριτές καταστάσεις τις οποίες έχει ορίσει ο σχεδιαστής του συστήματος, είτε εκφράζουν τις παρατηρήσεις που επηρεάζουν την εκτίμηση κατάστασης. Οι ακμές έχουν διαφορετική ερμηνεία, και εκφράζουν τη δυνατότητα μετάβασης μεταξύ των διάφορων καταστάσεων, με τις αριθμητικές τιμές να ποσοτικοποιούν το πόσο πιθανή είναι η εκάστοτε μετάβαση. Χρίζει αναφοράς ότι οι ακμές υποδηλώνουν τις σχέσεις εξάρτησης που επικρατούν στο HMM.



Σχήμα 3.2: Γενική μορφή ενός HMM.

Θα επεκταθούμε, συνοπτικά, στις σχέσεις εξάρτησης που διέπουν τη λειτουργία ενός HMM. Για παράδειγμα, στο Σχήμα 3.2, που αποτελεί τη γενική μορφή ενός HMM, οι ακμές υποδηλώνουν την εξάρτηση ότι η εκτίμηση της κατάστασης σε κάθε χρονικό πλαίσιο t εξαρτάται από την τρέχουσα παρατήρηση O_t , καθώς και από την εκτίμηση κατάστασης X_{t-1} του προηγούμενου χρονικού πλαισίου. Γενικότερα, οι κύριες ιδιότητες που διέπουν ένα HMM,

οι οποίες πηγάζουν αποκλειστικά από τη δομή και τον ορισμό του, εκφράζονται μέσω κάποιων σχέσεων ανεξαρτησίας. Συγκεκριμένα:

1. $O_t \perp O_{t-1} | X_t, X_{t-1}$
2. $O_t \perp X_{t-1} | X_t$
3. $X_t \perp X_{t-2} | X_{t-1}$

Οι ανεξαρτησίες αυτές προκύπτουν από την ευρύτερη θεωρία των Γραφικών Μοντέλων, και ειδικά από τους νόμους που διέπουν την ροή της πληροφορίας στα γραφικά μοντέλα ([37]), καθώς και από την γενικότερη θεώρηση ότι το δίκτυο ικανοποιεί τη *Μαρκοβιανή ιδιότητα*.

Ορισμός 3.2. Έστω ένα δίκτυο που μοντελοποιείται ως ένα HMM, και X_t η τυχαία μεταβλητή για την εκτίμηση της κατάστασης του δικτύου στο χρονικό πλαίσιο t . Θα λέμε ότι το δίκτυο εμφανίζει τη *Μαρκοβιανή ιδιότητα* (Markov property) εάν ισχύει:

$$P[X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0] = P[X_n = x_n | X_{n-1} = x_{n-1}] \quad (3.1)$$

δηλαδή η πιθανότητα η τυχαία μεταβλητή X_t να έχει μια συγκεκριμένη τιμή $x_t \in \mathcal{X}$ εξαρτάται μόνο από την προηγούμενη κατάσταση X_{t-1} , και όχι από την ακολουθία καταστάσεων που έχει προηγηθεί. Η *Μαρκοβιανή ιδιότητα* αναφέρεται συχνά και ως ιδιότητα έλλειψης μνήμης.

Έχοντας κάνει μια νύξη στην *Μαρκοβιανή ιδιότητα* και στις σχέσεις ανεξαρτησίας που διέπουν τα *Μαρκοβιανά μοντέλα*, θα αναλύσουμε, επιπλέον, τα απαραίτητα μεγέθη για τον πλήρη ορισμό ενός HMM. Συγκεκριμένα, απαραίτητα για τον ολοκληρωμένο ορισμό ενός κρυφού *Μαρκοβιανού μοντέλου* είναι:

- Οι πιθανότητες μετάβασης μεταξύ καταστάσεων. Αυτές εκφράζουν το πόσο πιθανή είναι η κάθε μετάβαση μεταξύ των υπάρχοντων καταστάσεων. Εάν υπάρχουν διαθέσιμα ακολουθιακά δεδομένα, τότε οι πιθανότητες δίνονται από τη σχέση:

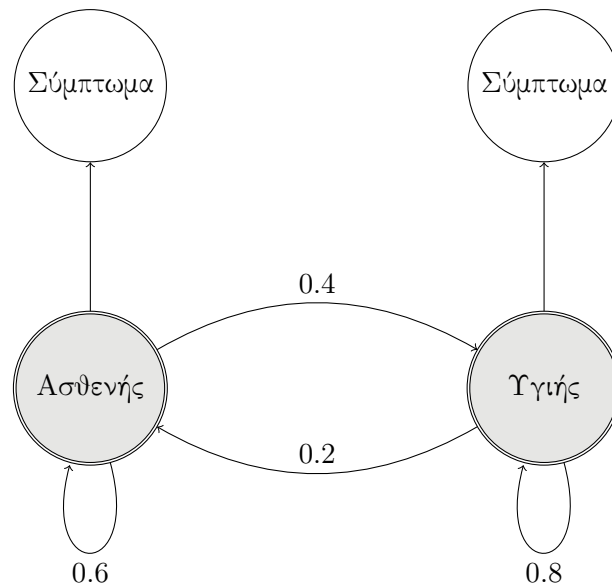
$$\alpha_{ij} = \frac{\sum_{t=1}^T P(\mathbf{O}, X_{t-1} = i, X_t = j)}{\sum_{t=1}^T P(\mathbf{O}, X_{t-1} = i)} \quad (3.2)$$

όπου η ποσότητα \mathbf{O} εκφράζει το σύνολο της ακολουθίας των παρατηρήσεων.

- Οι πιθανότητες παρατήρησης των δυνατών παρατηρήσεων σε κάθε κατάσταση. Αυτές εκφράζουν το πόσο πιθανή είναι η παρατήρηση του εκάστοτε συμβόλου εισόδου για κάθε ξεχωριστή κατάσταση. Υπό την ύπαρξη ακολουθιακών δεδομένων, οι τιμές προσδιορίζονται κάνοντας χρήση του αλγορίθμου Expectation Maximization (EM).
- Οι *a priori* πιθανότητες της κάθε κατάστασης. Αυτές εκφράζουν το κατά πόσο είναι πιθανό η διαφορετικές καταστάσεις να αποτελούν την αρχική κατάσταση. Υπό την ύπαρξη ακολουθιακών δεδομένων, αυτές δίνονται από τη σχέση:

$$\pi_i = \frac{\sum_{t=1}^T P(\mathbf{O}, X_0 = i)}{\sum_{t=1}^T P(\mathbf{O})} \quad (3.3)$$

Σε αυτό το σημείο θα εξετάσουμε ένα σύντομο παράδειγμα, παρουσιάζοντας τα προαναφερθείσα μεγέθη, για να κατανοηθεί από τον αναγνώστη λίγο καλύτερα η δομή των HMM.



Σχήμα 3.3: HMM για την έκφραση ασθένειας.

Συγκεκριμένα, θα εξετάσουμε ένα παράδειγμα όπου προσπαθούμε να εντοπίσουμε εάν ένας άνθρωπος είναι υγιής ή όχι.

Στο παραπάνω παράδειγμα, ας υποθέσουμε ότι η μεταβλητή Σύμπτωμα μπορεί να πάρει τρεις διακριτές τιμές: *συνάχι*, *πυρετός*, και *κανένα* (υποδηλώνοντας την έλλειψη κάποιου συμπτώματος). Για να έχει οριστεί πλήρως το σύστημα, χρειάζεται να προσδιορίσουμε τους πίνακες με τις πιθανότητες μετάβασης, πιθανότητες παρατήρησης, και των *a priori* πιθανοτήτων:

- Από το HMM του Σχήματος 3.3 μπορούμε να εξάγουμε τις πιθανότητες μετάβασης. Πράγματι, αυτές είναι:

Πιθανότητες μετάβασης μεταξύ καταστάσεων		
Κατάσταση	Ασθενής	Υγιής
Ασθενής	0.6	0.4
Υγιής	0.8	0.2

- Υποθέτουμε ότι οι πιθανότητες παρατήρησης του καθ' ενός από τα συμπτώματα στις καταστάσεις του HMM δίνεται από τον παρακάτω πίνακα:

Πιθανότητες έκφρασης συμπτωμάτων			
Κατάσταση	Κανένα	Συνάχι	Πυρετός
Ασθενής	0.1	0.4	0.5
Υγιής	0.7	0.2	0.1

- Υποθέτουμε ότι οι *a priori* πιθανότητες για το παράδειγμα δίνονται από τον παρακάτω πίνακα:

Apriori πιθανότητες		
Κατάσταση	Ασθενής	Υγιής
Πιθανότητα	0.3	0.7

3.3 Weighted Finite State Transducers

Τα Weighted Finite State Transducers (WFSTs) ([29], [32], [43]) αποτελούν μια ιδιαίτερη κατηγορία γραφικών μοντέλων, τόσο λόγω της διαφορετικότητας της μοντελοποίησής τους, όσο και λόγω των επιπλέον δυνατοτήτων τους. Όπως ίσχυε για όλες τις δομές που έχουμε εξετάσει μέχρι στιγμής, η γραφική τους αναπαράσταση γίνεται με τη χρήση κόμβων και ακμών. Επιπλέον, αποτελούν δομή οι οποία επιδέχεται εισόδους (όπως τα HMM), όμως, αντίθετα, με τα HMM παράγουν και έξοδο. Πέραν αυτής της διαφοράς, ο κύριος χαρακτήρας τους είναι ριζικά διαφορετικός από αυτόν των HMM. Πράγματι, η μοντελοποίηση ενός συστήματος με HMM αποσκοπεί στον προσδιορισμό της κατάστασης του συστήματος, και την εξέλιξή της στο χρόνο. Αντίθετα, τα WFSTs αποσκοπούν στην μετατροπή (*transduction*) μιας ακολουθίας εισόδου στην πιο πιθανή ακολουθία εξόδου. Αυτή η βασική διαφορά στον σκοπό και στην ερμηνεία τους υπογραμμίζεται και από τον διαφορετικό τρόπο μοντελοποίησής τους. Τα WFSTs πηγάζουν από την Επιστήμη των Υπολογιστών, και, συγκεκριμένα, αποτελούν επέκταση κάποιων άλλων βασικών δομών της Επιστήμης των Υπολογιστών, τα οποία είναι γνωστά ως automata. Τα WFSTs υλοποιούν δύο σημαντικές επεκτάσεις των automata. Συγκεκριμένα:

- Τα automata αποτελούν αποδοχείς (*acceptors*). Αυτό σημαίνει ότι δεν παράγουν καμία ακολουθία εξόδου. Αντίθετα, απλώς ελέγχουν εάν η εκάστοτε ακολουθία εισόδου μπορεί να απαντηθεί ως κάποια αλληλουχία καταστάσεων μέσα στο δίκτυο. Εάν αυτό είναι δυνατό, τότε αποκρίνονται θετικά. Σε διαφορετική περίπτωση, αποκρίνονται αρνητικά, υποδηλώνοντας την απόρριψη της ακολουθίας εισόδου.
- Τα automata αποτελούν αποδοχείς δίχως δυνατότητα βαρών. Αυτό σημαίνει ότι απλώς δέχονται ή απορρίπτουν την ακολουθία εισόδου, χωρίς να παρέχουν κάποια πληροφορία για την πιθανότητα, έστω και σχετική, αυτής της ακολουθίας.

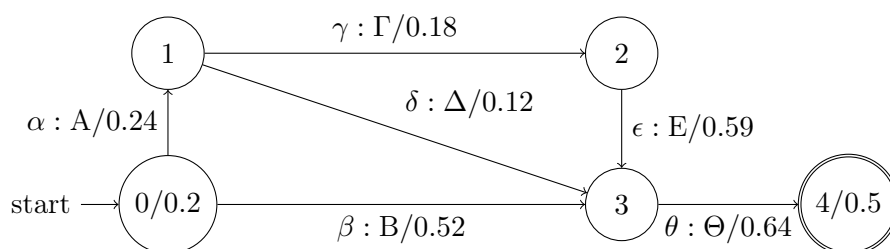
Λόγο της δομής τους, τα automata αποτελούν εξαιρετική δομή για τη μοντελοποίηση γλωσσών. Πράγματι, μοντελοποιώντας κατάλληλα τις πιθανές λέξεις που απαντώνται σε ένα λεξικό, ή τους ρητούς κανόνες που διέπουν μια γλώσσα, είναι δυνατό να αποφανθεί εάν μια συγκεκριμένη είσοδος αποτελεί λέξη της γλώσσας, ή εάν μια πρόταση αποτελεί συντακτικά και γραμματικά ορθή πρόταση της γλώσσας.

Θα προχωρήσουμε τώρα στον αυστηρό μαθηματικό ορισμό των WFSTs. Συγκεκριμένα, για τον πλήρη καθορισμό ενός WFST είναι αναγκαίος ο προσδιορισμός των παρακάτω ποσοτήτων:

- ένα σύνολο Q που εκφράζει τις καταστάσεις της δομής,
- ένα αλφάβητο Σ το οποίο εκφράζει τα σύμβολα εισόδου τα οποία απαντώνται στο WFST,
- ένα αλφάβητο Γ το οποίο εκφράζει τα σύμβολα εξόδου τα οποία απαντώνται στο WFST,
- ένα σύνολο $\Delta \subseteq Q \times (\Sigma \cup \{\epsilon\}) \times (\Gamma \cup \{\epsilon\}) \times Q$ το οποίο εκφράζει τις δυνατές μεταβάσεις μεταξύ των καταστάσεων του WFST,
- ένα σύνολο I το οποίο εκφράζει τις αρχικές καταστάσεις της δομής,

- ένα σύνολο F το οποίο εκφράζει τις τελικές καταστάσεις της δομής,
- ένα σύνολο λ το οποίο περιέχει τα βάρη εισόδου των αρχικών καταστάσεων της δομής,
- και ένα σύνολο ρ το οποίο περιέχει τα βάρη εκπομπής των τελικών καταστάσεων της δομής.

Είναι εξαιρετικά σπάνιος ο ορισμός ενός WFST μέσω της αυστηρής παράστασης των παραπάνω μεγεθών, καθώς είναι αρκετά, και αχρείαστα, σύνθετος. Είναι σημαντικά πιο σύνηθες να δίνονται κάποιοι βασικοί πίνακες (δηλαδή των βαρών μετάβασης, αρχικών και τελικών καταστάσεων) και μέσω αυτών να εξάγεται η πληροφορία για τα υπόλοιπα μεγέθη. Είναι συχνό, επίσης, να γίνεται η γραφική παράσταση ενός WFST, και να εξάγεται η σχετική πληροφορία από την σχηματική αναπαράσταση. Θα παρουσιάσουμε τώρα ένα WFST, και έπειτα θα εξάγουμε από αυτό τα μεγέθη του ορισμού, έτσι ώστε να εξοικειωθεί ο αναγνώστης με την εν λόγω δομή.



Σχήμα 3.4: WFST για μετατροπή πεζών σε κεφαλαία.

Από το Σχήμα 3.4 μπορούμε να εξάγουμε τα παρακάτω:

- Το σύνολο των καταστάσεων είναι εμφανές και ταυτίζεται με τις καταστάσεις που απαντώνται στο γράφημα.
- Το σύνολο των συμβόλων εισόδου/εξόδου αποτελείται από το σύνολο των συμβόλων που απαντώνται στις ακμές του Σχήματος 3.4.
- Το σύνολο των δυνατών μεταβάσεων υποδηλώνεται από τις ακμές του γραφήματος.
- Το σύνολο των αρχικών καταστάσεων υποδηλώνεται από το βέλος με την επισήμειωση *start*, και το σύνολο των τελικών καταστάσεων έχει επισημειωθεί με διπλό κύκλο.
- Τέλος, τα σύνολα των βαρών εισόδου και εκπομπής των αρχικών και τελικών καταστάσεων επισημειώνονται με μια διαγώνια γραμμή στις αντίστοιχες καταστάσεις.

Κεφάλαιο 4

Τροπικοί Αλγόριθμοι σε WFST

Στην Ενότητα 3.3 εισήγαμε την έννοια των WFST. Όπως είδαμε αυτές οι δομές δεδομένων έχουν σύνθετους μαθηματικούς ορισμούς, όμως είναι μεγίστης πρακτικής σημασίας σε πολλά σύγχρονα προβλήματα (χαρακτηριστικά στο Κεφάλαιο 3 αναφέραμε τις εφαρμογές τους σε ASR και NLP). Λόγω της μεγάλης πρακτικής εφαρμογής τους, υπήρχε ανέκαθεν η προσπάθεια να βελτιωθούν οι αλγόριθμοι οι οποίοι δρουν πάνω στα WFST ([42], [41], [43]).

Υπάρχει πληθώρα αλγορίθμων που εφαρμόζονται στα WFST. Ο πιο γνωστός και πολυμελετημένος είναι ο αλγόριθμος Viterbi ([20], [47], [52]), ο οποίος έχει ρίζες στο πεδίο των τηλεπικοινωνιών, και χρησιμοποιείται ώστε να εντοπιστεί η πιο πιθανή ακολουθία καταστάσεων στο WFST για μια ακολουθία διανυσμάτων παρατήρησης s_t . Ο αλγόριθμος Viterbi βασίζεται στις αρχές του δυναμικού προγραμματισμού ([13], [33], [35]). Οι αλγόριθμοι δυναμικού προγραμματισμού εγγυώνται τη βελτιστότητα των μετέπειτα επαναλήψεων τους δεδομένης της βελτιστότητας της τρέχουσας επανάληψης. Τέτοιοι αλγόριθμοι πάντα οδηγούν στην καλύτερη δυνατή λύση, αλλά είναι συχνά υπολογιστικά απαιτητικοί, καθώς ο χώρος των πιθανών λύσεων μπορεί να είναι εξαιρετικά μεγάλος. Στις εφαρμογές των WFST υπάρχει συχνά η απαίτηση για εύρεση πιθανών μονοπατιών σε πραγματικό χρόνο και έτσι είναι συχνό να χρησιμοποιείται μια παραλλαγή του αλγορίθμου Viterbi. Στην εν λόγω παραλλαγή σε κάθε βήμα αποκόπτονται κάποια από τα πιθανά μονοπάτια, θυσιάζοντας, πιθανώς, την βελτιστότητα στον βωμό της χαμηλότερης πολυπλοκότητας (οδηγώντας, συνεπώς, στον γρηγορότερο χρόνο εκτέλεσης).

Εκτός, όμως, από αλγορίθμους οι οποίοι αφήνουν πλήρως αναλλοίωτο το δίκτυο και απλώς ‘χρησιμοποιούν’ τα WFST για τον υπολογισμό των πιθανότερων μονοπατιών (γνωστού και ως decoding στον τομέα των τηλεπικοινωνιών, ή ως inference στον τομέα των πιθανοτικών γραφικών μοντέλων ([31], [37], [53]), υπάρχουν και αλγόριθμοι οι οποίοι αποσκοπούν στην αλλαγή τη δομής του WFST σε μια προσπάθεια βελτιστοποίησής του, διατηρώντας όμως όλη την αρχική πληροφορία που εκφραζόταν μέσω του WFST. Συνήθως, η εφαρμογή αυτών των αλγορίθμων αποσκοπεί στο να μειωθεί περαιτέρω ο χρόνος εκτέλεσης, έμμεσα ή άμεσα. Αλγόριθμοι που προσπαθούν άμεσα να βελτιώσουν τον χρόνο εκτέλεσης αποσκοπούν στο να μικρύνει το μέγεθος της δομής του WFST (δηλαδή να μειώσουν τον αριθμό των καταστάσεων και των μεταβάσεών του), και συνεπώς, καθώς έχει μικρότερο όγκο, να επιταχυνθεί άμεσα ο χρόνος εκτέλεσης. Από την άλλη, κάποιοι αλγόριθμοι δεν μικραίνουν τον όγκο της δομής δεδομένων, αλλά αντίθετα προσπαθούν να εισάγουν μεταβολές στα βάρη μεταξύ των κόμβων, οι οποίες θα οδηγήσουν σε γρηγορότερη εύρεση των βέλτιστων μονοπατιών. Αυτό επιτυγχάνεται μέσω της εφαρμογής της παραλλαγής του Viterbi με αποκοπή, όπου στο WFST με τα μεταβεβλημένα βάρη αποκόπτεται μεγαλύτερος αριθμός μονοπατιών στα αρχικά στάδια του αλγορίθμου, και έτσι έμμεσα βελτιώνεται ο χρόνος εκτέλεσης. Ένα εξαιρετο σύγγραμμα για τη μελέτη των αλγορίθμων που εφαρμόζονται στα WFST, με έμφαση την Αυτόματη Αναγνώριση Φωνής είναι το [29].

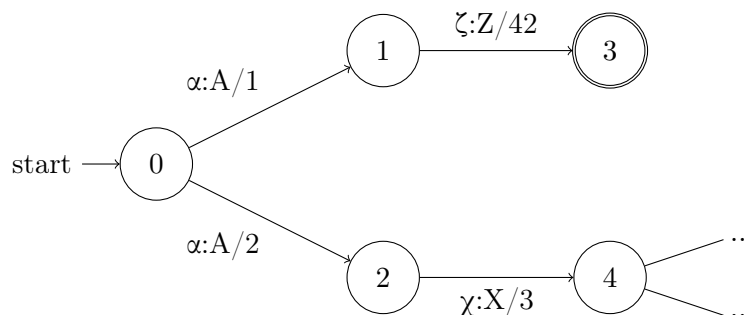
Στο υπόλοιπο του κεφαλαίου θα εξετάσουμε πιο προσεκτικά μερικούς από τους αλγορίθμους που εφαρμόζονται στα WFST, και θα τους μοντελοποιήσουμε στο πεδίο της τροπικής άλγεβρας. Μέσω αυτής της μοντελοποίησης αποσκοπούμε στο να αναδείξουμε την κομψότητα της συγκεκριμένης άλγεβρας στην έκφραση αυτών των αλγορίθμων, και παράλληλα, όπου είναι δυνατό, να εκμεταλλευτούμε ορισμένα από τα εργαλεία της τροπικής γεωμετρίας ώστε να εξάγουμε επιπλέον συμπεράσματα. Οι αλγόριθμοι οι οποίοι θα αναλύσουμε είναι:

- ο αλγόριθμος προώθησης βαρών,
- ο αλγόριθμος αφαίρεσης ε -μεταβάσεων,
- ο αλγόριθμος Viterbi,
- και ο αλγόριθμος Viterbi με αποκοπή.

Στα επερχόμενα χωρία θα εισάγουμε, αρχικά, τον εκάστοτε αλγόριθμο με την κλασική του έκφραση, συζητώντας συνοπτικά τη λειτουργία, τη χρησιμότητα του, καθώς και πιθανά λεπτά σημεία. Έπειτα θα αναλύσουμε τον αλγόριθμο με τη χρήση της τροπικής άλγεβρας (και, πιθανώς, της τροπικής γεωμετρίας), και τέλος θα παρουσιάσουμε κάποιο αριθμητικό παράδειγμα, για να υπογραμμίσουμε την κομψότητα της μοντελοποίησης, και να επαληθεύσουμε την ορθότητα των αποτελεσμάτων.

4.1 Αλγόριθμος προώθησης βαρών

Ο αλγόριθμος προώθησης βαρών αποτελεί έναν αλγόριθμο ο οποίος αποσκοπεί στην μεταβολή της δομής του WFST ώστε να βελτιώσει έμμεσα τον χρόνο που απαιτείται για την εκτέλεση του αλγορίθμου Viterbi με αποκοπή. Η λειτουργία του έχει ως σκοπό την μεταφορά βαρών (τόσο των βαρών μεταξύ των ακμών, αλλά και των βαρών εκπομπής των τελικών κόμβων) σε τοπολογικά αρχικότερες καταστάσεις του WFST, χωρίς όμως να αλλάξει η δομική πληροφορία που εμπεριέχεται σε αυτό, καθώς και χωρίς να αλλοιώσει τα βέλτιστα μονοπάτια. Αυτή η μεταβολή έχει ως αποτέλεσμα όταν εκτελείται ο αλγόριθμος Viterbi με αποκοπή, τα μονοπάτια να έχουν μεγαλύτερα κόστη στα αρχικά στάδια εκτέλεσης του αλγορίθμου. Έτσι, είναι πιο πιθανό να αποκοπούν μονοπάτια τα οποία σε αρχικό στάδιο φαίνεται να έχουν μεγάλη πιθανότητα να είναι τα βέλτιστα, αλλά όσο προχωράνε τα βήματα του αλγορίθμου αναδεικνύεται η μη βελτιστότητά τους.



Σχήμα 4.1: Παράδειγμα χρησιμότητας του αλγορίθμου προώθησης βαρών.

Στο παράδειγμα του Σχήματος 4.1 βλέπουμε ότι βρισκόμενοι στην κατάσταση 0 με σύμβολο εισόδου α μπορούμε να μεταβούμε είτε στην κατάσταση 1 είτε στην κατάσταση 2. Όμως,

το μοναδικό μονοπάτι που συνεχίζει από την κατάσταση 1 είναι αρκετά απίθανο καθώς έχει πολύ μεγάλο βάρος (για παράδειγμα μπορεί να αντιστοιχεί σε μια πολύ σπάνια λέξη). Επειδή η κατάσταση 3 είναι τελική (και είναι η μοναδική κατάσταση την οποία μπορούμε να καταλήξουμε από την κατάσταση 1), μπορούμε να ‘σπρώξουμε’ το βάρος της ακμής μεταξύ των καταστάσεων 1 και 3 στην ακμή μεταξύ των καταστάσεων 0 και 1. Έτσι, εάν τρέξουμε τον αλγόριθμο Viterbi με αποκοπή, βρισκόμενοι στην κατάσταση 0 με σύμβολο εισόδου a ερχόμαστε αντιμέτωποι με ένα μονοπάτι βάρους 43 και ένα μονοπάτι βάρους 2. Καθώς το βάρος του πρώτου μονοπατιού είναι υπερβολικά μεγάλο, το εν λόγω μονοπάτι θα αποκοπεί και δεν θα εξεταστεί ποτέ, γλιτώνοντας πολύτιμο υπολογιστικό χρόνο στον αλγόριθμο.

4.1.1 Ανάλυση

Παραδοσιακά, ο αλγόριθμος χωρίζεται σε δύο ρουτίνες: μια ρουτίνα η οποία υπολογίζει το δυναμικό κάθε κόμβου, που αποτελεί και τον κορμό του αλγορίθμου, και μια ρουτίνα η οποία αναλαμβάνει την ανανέωση των βαρών. Διαισθητικά, η πρώτη ρουτίνα υπολογίζει το βάρος που θα μπορούσε ‘εν δυνάμει’ να διαβεί τον κάθε κόμβο χωρίς να αλλοιωθεί η δομή του WFST, και έπειτα αυτό διαδίδεται προς τα πίσω. Αυστηρά, η ρουτίνα υπολογισμού των δυναμικών των κόμβων εκφράζεται σε ψευδοκώδικα ως εξής:

Algorithm 1 WFST-Potential(T)

```

1: for each  $q \in Q$  do
2:   if  $q \in F$  then
3:      $\mathbf{v}[q] \leftarrow \mathbf{r}[q] \leftarrow \rho[q]$ 
4:   else
5:      $\mathbf{v}[q] \leftarrow \mathbf{r}[q] \leftarrow \infty$ 
6:   end if
7: end for
8:  $S \leftarrow F$ 
9: while  $S \neq \emptyset$  do
10:   $q \leftarrow \text{Head}(S)$ 
11:   $\text{Dequeue}(S)$ 
12:   $R \leftarrow \mathbf{r}[q]$ 
13:   $\mathbf{r}[q] \leftarrow \infty$ 
14:  for each  $e \in E^{-1}[q]$  do
15:    if  $\mathbf{v}[p(e)] \neq \mathbf{v}[p(e)] \wedge (R + \mathbf{A}[e])$  then
16:       $\mathbf{v}[p(e)] \leftarrow \mathbf{v}[p(e)] \wedge (R + \mathbf{A}[e])$ 
17:       $\mathbf{r}[p(e)] \leftarrow \mathbf{v}[p(e)] \wedge (R + \mathbf{A}[e])$ 
18:      if  $p(e) \notin S$  then
19:         $\text{Enqueue}(S, p(e))$ 
20:      end if
21:    end if
22:  end for
23: end while
24: return  $V$ 

```

όπου με $p(e)$ δηλώνουμε την κατάσταση αναχώρησης της ακμής e και με $n(e)$ την κατάσταση άφιξης της ακμής e . Επιπλέον, με $E^{-1}[q]$ δηλώνουμε το σύνολο των ακμών που εισέρχονται στην κατάσταση q . Τέλος, χρησιμοποιούμε για απλότητα καταχρηστικά τον συμβολισμό $\mathbf{A}[e]$ έναντι του σωστότερου $\mathbf{A}_{p(e)n(e)}$.

Η ρουτίνα ανανέωσης των βαρών επαυξάνει τις τιμές του διανύσματος αρχικών βαρών με την τιμή του δυναμικού του εκάστοτε κόμβου, μειώνει τις τιμές του διανύσματος βαρών εκπομπής με την τιμή του δυναμικού του εκάστοτε κόμβου, και ανανεώνει τις τιμές των ακμών, μειώνοντας αυτές κατά την τιμή του δυναμικού του κόμβου αναχώρησης, και επαυξάνοντας κατά την τιμή του δυναμικού του κόμβου άφιξης.

Θα αφιερώσουμε λίγο χώρο στην απομυθοποίηση του παραπάνω αλγορίθμου, θα προσπαθήσουμε, δηλαδή, να εξηγήσουμε το πως λειτουργεί και πως, εν τέλει, επιτυγχάνει τον επιθυμητό σκοπό. Ο αλγόριθμος ξεκινά από τελικές καταστάσεις (οι οποίες είναι οι μόνες οι οποίες αρχικοποιούνται να έχουν πεπερασμένο δυναμικό στην γραμμή 3, μέσω του διανύσματος των βαρών εκπομπής). Ο λόγος γι' αυτό είναι ότι το δυναμικό του κάθε κόμβου περιορίζεται από το βάρος που μπορεί να διαδοθεί σε αυτόν τον κόμβο από τους μετέπειτα κόμβους. Για παράδειγμα, ένας κόμβος q μπορεί να έχει ένα μονοπάτι με βάρος 50 το οποίο να μπορεί να διαβιβαστεί προς τα πίσω. Εάν, όμως, τα υπόλοιπα μονοπάτια του κόμβου q δεν μπορούν να διαβιβάσουν και αυτά αντίστοιχο βάρος, τότε αναθέτοντας το δυναμικό του κόμβου q να είναι ίσο με 50 θα αλλάζαμε τη δομή του WFST, καθώς αυτό θα υποδήλωνε ότι όλα τα μονοπάτια που ξεκινούν από τον κόμβο q έχουν μεγάλο βάρος (και συνεπώς είναι απίθανα), κάτι το οποίο στο προκείμενο παράδειγμα δεν ισχύει.

Σε κάθε βήμα του βρόχου του αλγορίθμου (γραμμή 9), επιλέγεται μια κατάσταση q από τις προς εξέταση καταστάσεις και εξετάζονται οι ακμές που καταλήγουν σε αυτή την κατάσταση. Εάν το δυναμικό της προηγούμενης κατάστασης p που καταλήγει στην q είναι μεγαλύτερο του δυναμικού της κατάστασης q , προσαυξημένο με το βάρος της μεταξύ τους ακμής, τότε η τιμή του δυναμικού της κατάστασης p ανανεώνεται, και η κατάσταση προστίθεται στη λίστα των καταστάσεων προς εξέταση.

Το κομβικό σημείο εδώ είναι ότι κρατείται πάντα η ελάχιστη τιμή του δυναμικού που μπορεί να διαβεί από τον εκάστοτε κόμβο. Όπως εξηγήσαμε παραπάνω, αυτό αποτελεί κύρια απαίτηση, καθώς εξασφαλίζει ότι η συνολική δομή του WFST θα μείνει αναλλοίωτη. Αυτό επιτυγχάνεται βασιζόμενοι στο γεγονός ότι κρατώντας πάντα το ελάχιστο δυναμικό, όλες οι μετέπειτα καταστάσεις έχουν τουλάχιστον τόσο βάρος για να 'σπρώξουν' προς τις τοπολογικά πρότερες καταστάσεις.

Θα αναλύσουμε τώρα τον αλγόριθμο προώθησης βαρών χρησιμοποιώντας την τροπική άλγεβρα. Υποστηρίζουμε ότι ένα βήμα ανανέωσης (αναφερόμενοι σε μια επανάληψη του βρόχου της γραμμής 9) του Αλγορίθμου 1 μπορεί να γραφτεί στη μορφή:

$$\mathbf{v}_{i+1} = \mathbf{v}_i \wedge \mathbf{A} \boxplus \mathbf{v}_i, \quad (4.1)$$

όπου \mathbf{v}_i είναι το διάνυσμα των δυναμικών στην επανάληψη i , με $\mathbf{v}_0 = \boldsymbol{\rho}$, όπου $\boldsymbol{\rho}$ είναι το διάνυσμα βαρών εκπομπής.

Για να δώσουμε υπόσταση στον ισχυρισμό ας θεωρήσουμε αρχικά τον εσωτερικό βρόχο της γραμμής 9 του αλγορίθμου. Ο βρόγχος αυτός θα διατρέξει όλες τις ακμές οι οποίες εισέρχονται στον εξεταζόμενο κόμβο q , αλλάζοντας την τιμή δυναμικού του προηγούμενου κόμβου όταν η τιμή του δυναμικού του q προσαυξημένη από το αντίστοιχο στοιχείο της μεταξύ τους ακμής ξεπερνάει την τρέχουσα τιμή στο διάνυσμα δυναμικών. Η ποσότητα R του κλασικού αλγορίθμου (που προκύπτει από την εκάστοτε τιμή του διανύσματος \mathbf{r}) έχει πάντα την ίδια τιμή με αυτή του διανύσματος δυναμικού \mathbf{v} , όταν ο κόμβος q είναι υπό εξέταση. Αυτό συμβαίνει για τον εξής λόγο: οι τιμές των διανυσμάτων \mathbf{v} και \mathbf{r} ανανεώνονται συγχρονισμένα. Η μόνη στιγμή ασυμφωνίας των τιμών για μια δεδομένη κατάσταση q των δύο διανυσμάτων είναι αφού η κατάσταση εξεταστεί. Όμως, όταν εξεταστεί η κατάσταση παύει να έχει σημασία ο συγχρονισμός των δύο διανυσμάτων, καθώς η κατάσταση αφαιρέθηκε από τη λίστα. Η μοναδική προβληματική περίπτωση θα ήταν εάν η κατάσταση ξαναεισερχόταν στη λίστα, με τα

δύο διανύσματα σε ασυμφωνία. Κάτι τέτοιο όμως δεν είναι επικτό, καθώς εάν ξαναεισαχθεί η κατάσταση στη λίστα τότε θα ξαναεπιβληθεί η ομοφωνία των δύο διανυσμάτων (γραμμές 16-17). Η συμπιεσμένη μορφή της Εξίσωσης (4.1) αναθέτει σε κάθε στοιχείο j του διανύσματος \mathbf{v}_{i+1} το ελάχιστο μεταξύ της προηγούμενης τιμής του και της ποσότητας $(\mathbf{A} \boxplus \mathbf{v}_i)_j$, το οποίο αντιστοιχεί στην τιμή του δυναμικού της εξεταζόμενης κατάστασης q προσαυξημένη με το βάρος μετάβασης από τον κόμβο p στον κόμβο q .

Σε αυτό το σημείο ο παρατηρητικός αναγνώστης θα θέλει να εκφράσει ενστάσεις. Ενώ σε έναν βρόγχο του κλασικού αλγορίθμου κάθε κατάσταση p ελέγχεται το πολύ μια φορά (καθώς υπάρχει περίπτωση να μην ελεγχθεί καμία φορά εάν οι κόμβοι δεν ενώνονται με κάποια ακμή) για την ανανέωση του δυναμικού της σε κάθε εκτέλεση, στην προτεινόμενη γραφή του βήματος ανανέωσης της Εξίσωσης (4.1) κάθε κατάσταση p ελέγχεται n 'φορές' για την ανανέωση του δυναμικού της, λόγω του τροπικού πολλαπλασιασμού πινάκων (αφού υπολογίζεται το ελάχιστο μεταξύ n διαφορετικών τιμών για να ανατεθεί, τελικά, ως το δυναμικό της κατάστασης p). Κάτι τέτοιο φαντάζει κοστοφόρο και 'αχρείαστο', καθώς οι περισσότερες από αυτές τις καταστάσεις δεν είναι ακόμα ενεργές, δηλαδή δεν έχει γίνει η απαραίτητη οπισθοδρόμηση ώστε να έχουν πεπερασμένες τιμές στο διάνυσμα \mathbf{v}_i . Όμως, θα δείξουμε ότι κάτι τέτοιο, εν τέλει, δεν αποτελεί πρόβλημα. Πράγματι, εξετάζοντας την υπολογιστική πολυπλοκότητα, βλέπουμε ότι κατά την διάρκεια της εκτέλεσης του Αλγορίθμου 1 όλες οι καταστάσεις θα διέλθουν από τον βρόγχο της γραμμής 9 και κατά συνέπεια θα ήταν απαραίτητες n επαναλήψεις του βρόγχου. Επειδή οι κόμβοι διατρέχονται με βάση την τοπολογική διάταξη του γράφου (και θεωρούμε ότι η ουρά είναι *First In, First Out (FIFO)*), ο κάθε κόμβος θα εξεταστεί μια ακριβώς φορά. Συνεπώς, η υπολογιστική πολυπλοκότητα του κλασικού αλγορίθμου είναι $O(nm)$, όπου το n είναι το πλήθος των κόμβων και m είναι το πλήθος των ακμών, που μπορεί να είναι στην τάξη του $O(n^2)$. Συνεπώς, το άνω φράγμα της πολυπλοκότητας της χειρότερης περίπτωσης του κλασικού αλγορίθμου είναι $O(n^3)$.

Θα δείξουμε ότι η συνολική έκφραση του Αλγορίθμου 1 στην τροπική άλγεβρα έχει την ίδια πολυπλοκότητα χειρότερης περίπτωσης. Όπως αναφέραμε, η Εξίσωση (4.1) εκφράζει ένα μονάχα βήμα του αλγορίθμου. Το τελικό διάνυσμα δυναμικών προκύπτει από τη σχέση:

$$\mathbf{v}_N = \boldsymbol{\rho} \wedge \mathbf{A} \boxplus \boldsymbol{\rho} \wedge \mathbf{A} \boxplus (\mathbf{A} \boxplus \boldsymbol{\rho}) \wedge \dots \wedge \mathbf{A}^n \boxplus \boldsymbol{\rho} \wedge \dots \quad (4.2)$$

Για να εξηγήσουμε την παραπάνω εξάπλωση, ας εξετάσουμε την δεύτερη επανάληψη της συμπυκνωμένης μορφής:

$$\mathbf{v}_2 = \mathbf{v}_1 \wedge \mathbf{A} \boxplus \mathbf{v}_1$$

Όμως, $\mathbf{v}_1 = \mathbf{v}_0 \wedge \mathbf{A} \boxplus \mathbf{v}_0$, και $\mathbf{v}_0 = \boldsymbol{\rho}$, οπότε:

$$\mathbf{v}_2 = (\boldsymbol{\rho} \wedge \mathbf{A} \boxplus \boldsymbol{\rho}) \wedge \mathbf{A} \boxplus (\boldsymbol{\rho} \wedge \mathbf{A} \boxplus \boldsymbol{\rho}) = \boldsymbol{\rho} \wedge \mathbf{A} \boxplus \boldsymbol{\rho} \wedge \mathbf{A} \boxplus (\mathbf{A} \wedge \boldsymbol{\rho})$$

Αντικαθιστώντας με αντίστοιχο τρόπο και τις υπόλοιπες επαναλήψεις καταλήγουμε στην συμπυκνωμένη μορφή. Ο προσεκτικός αναγνώστης θα μπορεί να αναγνωρίσει στην Εξίσωση (4.2) τον πίνακα $\Delta(\mathbf{A})$ που ορίσαμε στο Κεφάλαιο 2. Πράγματι:

$$\mathbf{v}_N = (\mathbf{I} \wedge \mathbf{A} \wedge \mathbf{A}^2 \wedge \dots \wedge \mathbf{A}^n \wedge \dots) \boxplus \boldsymbol{\rho} = \Delta(\mathbf{A}) \boxplus \boldsymbol{\rho}. \quad (4.3)$$

Ισχυρισμός 4.2. Υποστηρίζουμε ότι ο υπολογισμός της Εξίσωσης (4.3):

$$\mathbf{v}_N = \Delta(\mathbf{A}) \boxplus \boldsymbol{\rho}$$

είναι πεπερασμένος, και επιπλέον $\Delta(\mathbf{A}) = \mathbf{I} \wedge \mathbf{A} \wedge \dots \wedge \mathbf{A}^{n-1}$.

Ο ισχυρισμός αυτός επαληθεύεται από το γεγονός ότι έχουμε υποθέσει ότι δεν υπάρχουν κύκλοι αρνητικού μήκους στον γράφο (μια συνήθης υπόθεση σε αυτά τα δίκτυα, καθώς τα βάρη προέρχονται από αρνητικούς λογαρίθμους πιθανοτήτων). Σε αυτή την περίπτωση, η υπολογιστική πολυπλοκότητα του παραπάνω είναι τετριμμένη: όπως δείξαμε στο Κεφάλαιο 2, ο πίνακας $\Delta(\mathbf{A})$ αποτελείται από το τροπικό άθροισμα των πινάκων $\Gamma(\mathbf{A})$, ο οποίος είναι ένας πίνακας ελαχίστων αποστάσεων και μπορεί να υπολογιστεί σε χρόνο ανάλογο του $O(n^3)$, και του τροπικού ταυτοτικού πίνακα.

Αφού έχουμε υπολογίσει το διάνυσμα δυναμικών \mathbf{v}_N μπορούμε να προχωρήσουμε στην ανανέωση των βαρών του WFST. Για την ανανέωση θα χρειαστεί να ορίσουμε κάποια επιπλέον διανύσματα, καθώς και κάποιους πίνακες, τα οποία θα μας επιτρέψουν να εκφράσουμε όλους τους υπολογισμούς της ανανέωσης εξ' ολοκλήρου στο συμβολισμό της τροπικής άλγεβρας. Συγκεκριμένα, χρειαζόμαστε το διάνυσμα δυναμικών \mathbf{v}_N που υπολογίσαμε, καθώς και το αρνητικό διάνυσμα δυναμικών \mathbf{v}_{-N} . Επιπλέον, χρειαζόμαστε 4 διαγώνιους πίνακες:

1. Τον διαγώνιο πίνακα \mathbf{A} των αρχικών βαρών.
2. Τον διαγώνιο πίνακα \mathbf{V}_N του διανύσματος δυναμικών \mathbf{v}_N .
3. Τον διαγώνιο πίνακα \mathbf{V}_{-N} του αρνητικού διανύσματος δυναμικών \mathbf{v}_{-N} .
4. Τον διαγώνιο πίνακα \mathbf{P} του διανύσματος εκπομπής.

Έχοντας πλέον ορίσει όλους τους απαραίτητους πίνακες και διανύσματα, οι ανανεωμένοι πίνακες του μοντέλου, όπου έχει εφαρμοστεί ο αλγόριθμος προώθησης βαρών, είναι:

- $\mathbf{A}' = \mathbf{A} \boxplus \mathbf{v}_N$, αφού στον αλγόριθμο ανανέωσης κάθε αρχικό βάρος είναι *προσαυξημένο* κατά την τιμή του δυναμικού στον αντίστοιχο κόμβο.
- $\mathbf{P}' = \mathbf{P} \boxplus \mathbf{v}_{-N}$, αφού στον αλγόριθμο ανανέωσης κάθε βάρος εκπομπής είναι *μειωμένο* κατά την τιμή του δυναμικού στον αντίστοιχο κόμβο.
- $\mathbf{A}' = \mathbf{V}_{-N} \boxplus \mathbf{A} \boxplus \mathbf{V}_N$, αφού στον αλγόριθμο ανανέωσης κάθε βάρος είναι *μειωμένο* κατά την τιμή του δυναμικού στον εισερχόμενο κόμβο και *προσαυξημένο* κατά την τιμή του δυναμικού στον εξερχόμενο κόμβο.

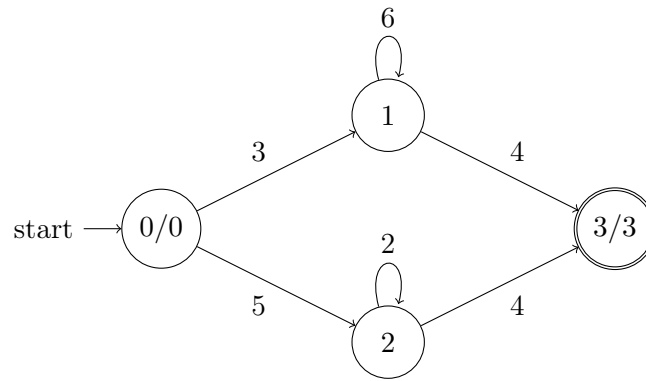
Ο συγκεκριμένος αλγόριθμος αποτελεί ένα από τα καλύτερα παραδείγματα για να ξεκινήσουμε την ανάλυσή μας, καθώς κάνει εμφανές ένα από τα κύρια εναύσματα μας για τη χρήση της τροπικής άλγεβρας: η μοντελοποίησή του υπογραμμίζει το πόσο πιο κομψή και κατανοητή είναι η έκφραση του στην τροπική άλγεβρα, συγκριτικά με τον γραμμικό αλγόριθμο.

Παράδειγμα 4.1. Σε αυτή την ενότητα θα εξετάσουμε ένα παράδειγμα για να υπογραμμίσουμε τις διαφορές μεταξύ του Αλγορίθμου 1 για την προώθησης βαρών, και της προτεινόμενης μοντελοποίησης της Εξίσωσης (4.3). Ας θεωρήσουμε το παρακάτω στοιχειώδες δίκτυο, το οποίο έχει μόνο μια τελική κατάσταση. Ο πίνακας μεταβάσεων για το συγκεκριμένο δίκτυο του Σχήματος 4.2 είναι:

$$\mathbf{A} = \begin{bmatrix} \infty & 3 & 5 & \infty \\ \infty & 6 & \infty & 4 \\ \infty & \infty & 2 & 4 \\ \infty & \infty & \infty & \infty \end{bmatrix}$$

Επιπλέον, έστω ότι το διάνυσμα των αρχικών βαρών και το διάνυσμα βαρών εκπομπής έχουν τις εξής μορφές, αντίστοιχα:

$$\boldsymbol{\lambda} = \begin{bmatrix} 0 \\ \infty \\ \infty \\ \infty \end{bmatrix}, \quad \boldsymbol{\rho} = \begin{bmatrix} \infty \\ \infty \\ \infty \\ 3 \end{bmatrix}$$



Σχήμα 4.2: Δίκτυο για το παράδειγμα του αλγορίθμου προώθησης βαρών.

Ως ένα πρώτο βήμα θα εκτελέσουμε αναλυτικά τα βήματα του Αλγορίθμου 1, όπως αυτός περιγράφεται από τους Horii και Nakamura στο [29]. Αρχικοποιούμε τα διανύσματα \mathbf{v} και \mathbf{r} (γραμμή 3):

$$\mathbf{v} = \begin{bmatrix} \infty \\ \infty \\ \infty \\ 3 \end{bmatrix}, \quad \mathbf{r} = \begin{bmatrix} \infty \\ \infty \\ \infty \\ 3 \end{bmatrix}$$

Η ουρά S αρχικοποιείται με την τελική κατάσταση 3, δηλαδή $S = \{3\}$. Έπειτα, ο αλγόριθμος ξεκινά τις εκτελέσεις του εσωτερικού βρόγχου της γραμμής 9:

- Η κατάσταση $q = 3$ εξάγεται από την ουρά S :

$R \leftarrow \mathbf{r}[q] = 3$, στο $\mathbf{r}[q]$ εκχωρείται η τιμή ∞ , και το δυναμικό για τις δύο εισερχόμενες καταστάσεις γίνεται:

- $\mathbf{v}[1] = 7, \mathbf{r}[1] = 7$
- $\mathbf{v}[2] = 7, \mathbf{r}[2] = 7$

Τέλος, οι δύο καταστάσεις 1 και 2 εισέρχονται στην ουρά και η κατάσταση της ουράς πλέον είναι $S = \{1, 2\}$.

- Η κατάσταση $q = 1$ εξάγεται από την ουρά S :

$R \leftarrow \mathbf{r}[q] = 7$, στο $\mathbf{r}[q]$ εκχωρείται η τιμή ∞ , και το δυναμικό για τις δύο εισερχόμενες καταστάσεις γίνεται:

- $\mathbf{v}[1] = 7$ αναλλοίωτο
- $\mathbf{v}[0] = 10, \mathbf{r}[0] = 10$

Τέλος, η κατάσταση 0 εισέρχεται στην ουρά και η κατάσταση της ουράς πλέον είναι $S = \{2, 0\}$.

- Η κατάσταση $q = 2$ εξάγεται από την ουρά S :

$R \leftarrow \mathbf{r}[q] = 7$, στο $\mathbf{r}[q]$ εκχωρείται η τιμή ∞ , και το δυναμικό για τις δύο εισερχόμενες καταστάσεις γίνεται:

- $\mathbf{v}[2] = 7$ αναλλοίωτο

– $\mathbf{v}[0] = 10$ αναλλοίωτο

Η κατάσταση της ουράς πλέον είναι $S = \{0\}$.

- Η κατάσταση $q = 0$ εξάγεται από την ουρά S :

Δεν υπάρχουν εισερχόμενες καταστάσεις και η ουρά είναι κενή, επομένως ο αλγόριθμος τερματίζει.

Έχοντας ολοκληρωθεί η εκτέλεση του Αλγορίθμου 1, η τελική μορφή του διανύσματος \mathbf{v} είναι:

$$\mathbf{v} = \begin{bmatrix} 10 \\ 7 \\ 7 \\ 3 \end{bmatrix}$$

Προχωρώντας στο βήμα της ανανέωσης των μεγθών της δομής δεδομένων, τα νέα διανύσματα αρχικών βαρών, μεταβάσεων, και βαρών εκπομπής παίρνουν τις εξής μορφές:

$$\lambda' = \begin{bmatrix} 10 \\ \infty \\ \infty \\ \infty \end{bmatrix}, \quad \mathbf{A}' = \begin{bmatrix} \infty & 0 & 2 & \infty \\ \infty & 6 & \infty & 0 \\ \infty & \infty & 2 & 0 \\ \infty & \infty & \infty & \infty \end{bmatrix}, \quad \rho' = \begin{bmatrix} \infty \\ \infty \\ \infty \\ 0 \end{bmatrix}$$

Ας επαναλάβουμε, τώρα, το προηγούμενο παράδειγμα, χρησιμοποιώντας την προτεινόμενη συμπυκνωμένη μορφή του αλγορίθμου της Εξίσωσης (4.3). Αρχικά, ας υπολογίσουμε τον πίνακα $\Delta(\mathbf{A})$. Θα χρειαστεί πρώτα να υπολογίσουμε τους πίνακες \mathbf{A}^2 , και \mathbf{A}^3 . Έχουμε:

$$\mathbf{A} = \begin{bmatrix} \infty & 3 & 5 & \infty \\ \infty & 6 & \infty & 4 \\ \infty & \infty & 2 & 4 \\ \infty & \infty & \infty & \infty \end{bmatrix}, \quad \mathbf{A}^2 = \begin{bmatrix} \infty & 9 & 7 & 7 \\ \infty & 12 & \infty & 10 \\ \infty & \infty & 4 & 6 \\ \infty & \infty & \infty & \infty \end{bmatrix}, \quad \mathbf{A}^3 = \begin{bmatrix} \infty & 15 & 9 & 11 \\ \infty & 18 & \infty & 16 \\ \infty & \infty & 6 & 8 \\ \infty & \infty & \infty & \infty \end{bmatrix}$$

Επομένως, ο πίνακας $\Delta(\mathbf{A})$ είναι:

$$\Delta(\mathbf{A}) = (\mathbf{I} \wedge \mathbf{A} \wedge \mathbf{A}^2 \wedge \mathbf{A}^3) = \begin{bmatrix} 0 & 3 & 5 & 7 \\ \infty & 0 & \infty & 4 \\ \infty & \infty & 0 & 4 \\ \infty & \infty & \infty & 0 \end{bmatrix}$$

Συνεπώς, το τελικό διάνυσμα δυναμικών παίρνει την μορφή:

$$\mathbf{v}_N = \Delta(\mathbf{A}) \boxplus \rho = \begin{bmatrix} 10 \\ 7 \\ 7 \\ 3 \end{bmatrix}, \quad \mathbf{v}_{-N} = \begin{bmatrix} -10 \\ -7 \\ -7 \\ -3 \end{bmatrix}$$

Παρατηρούμε ότι, όπως ήταν αναμενόμενο, το τελικό διάνυσμα δυναμικών είναι πανομοιότυπο με το διάνυσμα το οποίο παράχθηκε από την εκτέλεση του Αλγορίθμου 1. Τέλος, ας ορίσουμε

τους πίνακες οι οποίοι είναι απαραίτητοι για να κάνουμε την ανανέωση των μεγεθών:

$$\mathbf{A} = \begin{bmatrix} 0 & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty \end{bmatrix}, \quad \mathbf{V}_N = \begin{bmatrix} 10 & \infty & \infty & \infty \\ \infty & 7 & \infty & \infty \\ \infty & \infty & 7 & \infty \\ \infty & \infty & \infty & 3 \end{bmatrix},$$

$$\mathbf{V}_{-N} = \begin{bmatrix} -10 & \infty & \infty & \infty \\ \infty & -7 & \infty & \infty \\ \infty & \infty & -7 & \infty \\ \infty & \infty & \infty & -3 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & 3 \end{bmatrix}$$

Τότε, τα ανανεωμένα διανύσματα των αρχικών βαρών, βαρών μετάβασης, και βαρών εκπομπής είναι:

$$\boldsymbol{\lambda}' = \mathbf{A} \boxplus \mathbf{v}_N = \begin{bmatrix} 10 \\ \infty \\ \infty \\ \infty \end{bmatrix}, \quad \boldsymbol{\rho}' = \mathbf{P} \boxplus \mathbf{v}_{-N} = \begin{bmatrix} \infty \\ \infty \\ \infty \\ 0 \end{bmatrix},$$

$$\mathbf{A}' = \mathbf{V}_{-N} \boxplus \mathbf{A} \boxplus \mathbf{V}_N = \begin{bmatrix} \infty & 0 & 2 & \infty \\ \infty & 6 & \infty & 0 \\ \infty & \infty & 2 & 0 \\ \infty & \infty & \infty & \infty \end{bmatrix}$$

Και πάλι παρατηρούμε ότι τα αποτελέσματα ταυτίζονται με αυτά που προέκυψαν από την εκτέλεση του Αλγόριθμου 1. Δεν μπορούμε παρά να εκτιμήσουμε πόσο πιο ευθύς είναι οι υπολογισμοί, και πόσο πιο σίγουροι είμαστε για το αποτέλεσμα συγκριτικά με την εφαρμογή του κλασικού αλγόριθμου. Ο υπολογισμός του πίνακα $\Delta(\mathbf{A})$ αποτελεί το μόνο μελανό σημείο της προσέγγισης, εννοώντας ότι είναι υπολογιστικά απαιτητικό, αλλά όπως εξηγήσαμε στην Ενότητα 4.1.1 η πολυπλοκότητες χειρότερης περίπτωσης του Αλγόριθμου 1 και της συμπυκνωμένης μορφής της Εξίσωσης (4.3) ταυτίζονται.

4.2 Αλγόριθμος αφαίρεσης ε -μεταβάσεων

Ο αλγόριθμος αφαίρεσης ε -μεταβάσεων αποτελεί έναν αλγόριθμο που αποσκοπεί στην μεταβολή της δομής του WFST ώστε να βελτιώσει άμεσα τον χρόνο που απαιτείται για την εκτέλεση του αλγορίθμου Viterbi, μέσω της μείωσης του αριθμού των καταστάσεων και των ακμών του δικτύου. Η λειτουργία του έχει ως σκοπό την αφαίρεση των μεταβάσεων μεταξύ καταστάσεων οι οποίες γίνονται με το σύμβολο ε (δηλαδή μεταβάσεις χωρίς σύμβολο εισόδου ή εξόδου). Τέτοιες μεταβάσεις προκύπτουν από κάποιους άλλους αλγορίθμους των WFSTs (συγκεκριμένα προκύπτουν συνήθως από τον αλγόριθμο της σύνθεσης δύο WFSTs). Αυτό συμβαίνει επειδή συχνά όταν άνθρωποι σχεδιάζουν WFSTs, τα σχεδιάζουν με τον τρόπο που είναι πιο άμεσα κατανοητά από τους ίδιους, χωρίς όμως αυτό να εγγυάται καμία βελτιστότητα για τη δομή τους. Η αφαίρεση αυτών των μεταβάσεων που προκύπτουν, και η μετέπειτα αφαίρεση των μη προσβάσιμων καταστάσεων, μπορούν να οδηγήσουν σε σημαντική μείωση του όγκου του WFST, οδηγώντας σε μεγάλη βελτίωση στον χρόνο εκτέλεσης.

4.2.1 Ανάλυση

Ο Αλγόριθμος 2 παρουσιάζει μια υλοποίηση του αλγορίθμου αφαίρεσης ε -μεταβάσεων. Σε ένα διαισθητικό επίπεδο, ο αλγόριθμος εντοπίζει όλα τα μονοπάτια τα οποία ξεκινούν από

κάποια κατάσταση αναχώρησης p και καταλήγουν σε κάποια κατάσταση άφιξης q αποκλειστικά με ε -μεταβάσεις, αφαιρεί αυτά τα μονοπάτια, και προσθέτει μια καινούργια, μη ε -μετάβαση έτσι ώστε το προκύπτον WFST να είναι ισοδύναμο με το αρχικό.

Algorithm 2 WFST-EpsilonRemoval(T)

```

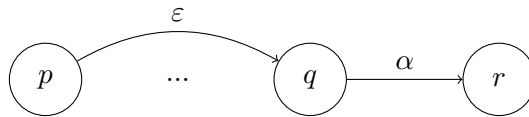
1: for each  $p \in Q$  do
2:    $E' \leftarrow E' \uplus \{e \in E[p] \mid (i[e], o[e]) \neq (\varepsilon, \varepsilon)\}$ 
3:   for each  $(q, w') \in \text{EpsilonClosure}(p)$  do
4:      $E' \leftarrow E' \uplus \{(p, x, y, w' + w, r) \mid (q, x, y, w, r) \in E[q], (x, y) \neq (\varepsilon, \varepsilon)\}$ 
5:     if  $p \in F$  then
6:        $F' \leftarrow F' \cup \{p\}$ 
7:        $\rho'(p) \leftarrow \rho(p)$ 
8:     end if
9:     if  $q \in F$  then
10:      if  $p \notin F$  then
11:         $F' \leftarrow F' \cup \{p\}$ 
12:      end if
13:       $\rho'(p) \leftarrow \rho'(p) \wedge (w' + \rho(q))$ 
14:    end if
15:  end for
16: end for
17: return  $T' = (\Sigma, \Delta, Q, I, F', E', \lambda, \rho')$ 

```

Θα εξετάσουμε συνοπτικά πως επιτυγχάνει τα παραπάνω ο Αλγόριθμος 2. Σε ένα πρώτο βήμα, ο αλγόριθμος αρχικοποιεί το νέο σύνολο ακμών E' να εμπεριέχει μόνο τις μεταβάσεις που ξεκινούν από μια κατάσταση p οι οποίες δεν είναι ε -μεταβάσεις. Έπειτα εξετάζεται το σύνολο της ε -κλειστότητας της κατάστασης p .

Ορισμός 4.3. Η ε -κλειστότητα μιας κατάστασης p είναι το σύνολο των καταστάσεων οι οποίες είναι προσβάσιμες από την p αποκλειστικά με ε -μεταβάσεις.

Για κάθε ζεύγος (q, w') στην ε -κλειστότητα της κατάστασης p (όπου q είναι η κατάσταση άφιξης και w' το βάρος του μονοπατιού) προστίθεται μια καινούργια ακμή στο νέο σύνολο E' . Η ακμή αυτή περιγράφεται στην γραμμή 4, την οποία θα εξηγήσουμε πιο προσεκτικά. Η ε -κλειστότητα περιέχει τις καταστάσεις q που είναι προσβάσιμες αποκλειστικά μέσω μονοπατιών ε -μεταβάσεων. Όμως, για να αφαιρεθούν οι ε -μεταβάσεις δεν μας ενδιαφέρει η μετάβαση στην κατάσταση q , αλλά μας ενδιαφέρουν μεταβάσεις από την q , οι οποίες δεν είναι και αυτές ε -μεταβάσεις.



Σχήμα 4.3: Βοηθητική επεξήγηση για τον Αλγόριθμο 2.

Στο παράδειγμα του Σχήματος 4.3 βλέπουμε ότι η κατάσταση q ανήκει στην ε -κλειστότητα της κατάστασης p , αλλά η προσθήκη μιας μετάβασης από την p στην q θα ήταν και πάλι ε -μετάβαση. Καταλαβαίνουμε, λοιπόν, ότι ενδιαφέρον παρουσιάζουν για τον αλγόριθμο οι μετέπειτα μεταβάσεις από την κατάσταση q σε κάποια κατάσταση r , οι οποίες δεν αποτελούν

ε -μεταβάσεις. Έτσι, η μετάβαση που προστίθεται στο νέο σύνολο ακμών E' στη γραμμή 4 απλώς εκφράζει αυτή τη νέα ακμή από την κατάσταση p στην κατάσταση r , με μεταβεβλημένο το βάρος ώστε η δομή του WFST να παραμένει αναλλοίωτη.

Θα μοντελοποιήσουμε τώρα τον Αλγόριθμο 2 για την αφαίρεση ε -μεταβάσεων στο πεδίο της τροπικής άλγεβρας. Αρχικά, θα χρειαστεί να ορίσουμε τρεις πίνακες, οι οποίοι χρησιμοποιούνται στα WFST, που θα μας επιτρέψουν να πορευτούμε στην ανάλυσή μας. Η δομή αυτών των πινάκων είναι ίδια, δηλαδή έχουν μη-άπειρα στοιχεία ακριβώς στις ίδιες θέσεις, και το μόνο που αλλάζει είναι οι συγκεκριμένες τιμές των πεπερασμένων αυτών στοιχείων. Συγκεκριμένα, οι πίνακες αυτοί είναι:

- ο πίνακας \mathbf{A} , που είναι ο γνωστός, πλέον, πίνακας των βαρών μετάβασης μεταξύ των καταστάσεων.
- ο πίνακας \mathbf{I} , που είναι ο πίνακας των συμβόλων εισόδου για τις μεταβάσεις (ο οποίος δεν πρέπει να μπερδευτεί με τον ταυτοτικό πίνακα διαστάσεων $n \times n$ της τροπικής άλγεβρας I_n).
- ο πίνακας \mathbf{O} , που είναι ο πίνακας των συμβόλων εξόδου για τις μεταβάσεις.

Πριν προχωρήσουμε στην ανάλυση, θα πρέπει να κάνουμε μερικές παραδοχές, έτσι ώστε να είναι δυνατή η μοντελοποίηση του Αλγορίθμου 2 για την αφαίρεση ε -μεταβάσεων. Κάποιες από αυτές τις παραδοχές ενδέχεται να μην ισχύουν σε ένα γενικό WFST, και θα συζητήσουμε τα προβλήματα που προκύπτουν, και πώς αυτά αντιμετωπίζονται στα WFST που χρησιμοποιούνται στις εφαρμογές του ενδιαφέροντός μας. Συγκεκριμένα:

- Οι μόνες μεταβάσεις οι οποίες θεωρούνται ως ε -μεταβάσεις στο πλαίσιο του αλγορίθμου (και συνεπώς οι μόνες μεταβάσεις οι οποίες θα αφαιρεθούν) είναι αυτές όπου και τα δύο σύμβολα, εισόδου και εξόδου (δηλαδή τα στοιχεία των πινάκων \mathbf{I} και \mathbf{O}), είναι ε . Μια τέτοια παραδοχή μπορεί να φαντάζει αρκετά περιοριστική, όμως είναι μια πολύ συνηθισμένη παραδοχή στο πλαίσιο των αλγορίθμων σε WFST, και μια παραδοχή την οποία την θεωρούν πληθώρα συγγραφέων του τομέα ([29], [42], [41]). Συχνά είναι αδύνατον να μην υπάρχουν καθόλου σύμβολα ε σε ένα πρακτικό WFST, και στην ουσία οι αλγόριθμοι για αφαίρεση ε -μεταβάσεων αποσκοπούν στην ελαχιστοποίησή τους, παρά στην ολική εξάλειψή τους. Στο πνεύμα αυτής της παραδοχής υποθέτουμε ότι έχει εφαρμοστεί και κάποιος αλγόριθμος συγχρονισμού (όπως αναφέρεται στο [29]), έτσι ώστε τα σύμβολα ε μεταξύ των δύο πινάκων \mathbf{I} και \mathbf{O} να προκύπτουν κατά το ει δυνατόν με τη μέγιστη συμφωνία, επιτρέποντας την επιτυχή αφαίρεσή τους.
- Στο τελικό WFST (δηλαδή αυτό που προκύπτει μετά την εφαρμογή του Αλγορίθμου 2 για την αφαίρεση ε -μεταβάσεων) δεν υπάρχουν πολλαπλές μεταβάσεις από μια προηγούμενη σε μια επόμενη κατάσταση. Επιτρέπουμε μονάχα μια μοναδική μετάβαση, ανεξάρτητα με το αν έχουν διαφορετικά σύμβολα εισόδου/εξόδου, ή/και βάρη. Αυτή η θεώρηση μπορεί να μοιάζει αυθαίρετη, αλλά είναι μια θεώρηση που έχουμε κάνει έμμεσα καθ' όλη τη διάρκεια του παρόντος κειμένου, αφού χρησιμοποιούμε πίνακες, οι οποίοι για κάθε ζεύγος κόμβων (i, j) επιδέχονται ένα μοναδικό στοιχείο. Η υπόθεση αυτή αποτελεί σίγουρα έναν ισχυρότερο περιορισμό από ότι ο προηγούμενος, και σε αυτό το σημείο μπορεί να παραβιαστεί σημαντικά ευκολότερα (από τη στιγμή που εμπλουτίζουμε τη δομή μας με περαιτέρω μεταβάσεις οι οποίες δεν ήταν παρούσες στο αρχικό WFST). Σε μια πρώτη όψη, αυτή η παραδοχή μπορεί να φαντάζει καταστροφική ως προς την εκφραστικότητα του μοντέλου, και αποτελεί ένα εμπόδιο το οποίο δεν μπορούμε να ξεπεράσουμε στη θεωρητική ανάλυση. Παρ' όλα αυτά, ο αναγνώστης δεν χρειάζεται να

θορυβείται: αυτή η περίπτωση αποτελεί μια από τις σπάνιες όπου η πρακτική εφαρμογή των ιδεών έρχεται να επιλύσει τα προβλήματα της θεωρίας. Ακόμα και μετά την αφαίρεση των ε -μεταβάσεων, οι καταστάσεις διατηρούν μια αφηρημένη αντιστοιχία με πραγματικές αναγνωρισμένες λέξεις. Η κάθε κατάσταση είναι προσβάσιμη μόνο από κάποια συγκεκριμένη ακολουθία συμβόλων εισόδου, και εκφράζει τη λέξη που προκύπτει από το το συνδυασμό αυτών των συμβόλων. Καταλαβαίνουμε ότι τότε πολλαπλές μεταβάσεις μεταξύ μιας κατάστασης i και μιας κατάστασης j δεν θα έπρεπε να είναι δυνατές, καθώς κάτι τέτοιο θα υπονοούσε ότι η κατάσταση j αντιστοιχεί σε δύο διαφορετικές αναγνωρισμένες λέξεις.

Έχοντας κάνει μια νύξη στις παραδοχές, μπορούμε πλέον να προχωρήσουμε στην ανάλυση. Θα χρειαστεί να ορίσουμε έναν ακόμα πίνακα που δεν χρησιμοποιείται παραδοσιακά στα WFST, ο οποίος θα μας επιτρέψει να υπολογίσουμε εύκολα την ε -κλειστότητα όλων των κόμβων του γράφου. Έστω ο πίνακας \mathbf{E} , όπου:

$$E_{ij} = \begin{cases} \alpha_{ij}, & \text{αν } I_{ij} = O_{ij} = \varepsilon \\ \infty, & \text{διαφορετικά} \end{cases} \quad (4.4)$$

Ισχυρισμός 4.3. Έστω \mathbf{E} ο πίνακας της Εξίσωσης (4.4). Τότε, ο πίνακας

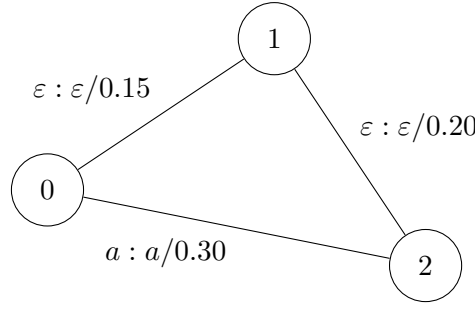
$$\Gamma(\mathbf{E}) = \mathbf{E} \wedge \mathbf{E}^2 \wedge \dots \wedge \mathbf{E}^n \wedge \dots \quad (4.5)$$

είναι πεπερασμένος και ίσος με $\Gamma(\mathbf{E}) = \mathbf{E} \wedge \mathbf{E}^2 \wedge \dots \wedge \mathbf{E}^{n-1}$, και επιπλέον εκφράζει την ε -κλειστότητα για όλες τις καταστάσεις του WFST.

Ο ισχυρισμός επαληθεύεται από το γεγονός ότι ο πίνακας \mathbf{E} περιέχει μόνο τις ε -μεταβάσεις. Συγκεκριμένα, ο αρχικός ισχυρισμός ότι ο πίνακας $\Gamma(\mathbf{E})$ είναι πεπερασμένος προκύπτει από την γενική παραδοχή που κάναμε, ότι δεν υπάρχουν κύκλοι αρνητικού μήκους. Έπειτα, το ότι ο πίνακας $\Gamma(\mathbf{E})$ εκφράζει την ε -κλειστότητα για όλες τις καταστάσεις προκύπτει αν αναλογιστούμε ότι ουσιαστικά ο πίνακας Γ στο πλαίσιο της τροπικής άλγεβρας υπολογίζει ελάχιστες αποστάσεις. Συνεπώς, εν προκειμένω ο πίνακας $\Gamma(\mathbf{E})$ υπολογίζει τις ελάχιστες αποστάσεις των κόμβων του δικτύου μόνο με ε -μεταβάσεις, που είναι ακριβώς ο ορισμός για την ε -κλειστότητα του Ορισμού 4.3.

Ο λόγος ο οποίος καθιστά αναγκαίο τον ορισμό του πίνακα \mathbf{E} είναι διττός: αφ' ενός, ο Αλγόριθμος 2 για την αφαίρεση ε -μεταβάσεων θεωρεί ως ε -μετάβαση μια μετάβαση όπου και το σύμβολο εισόδου και το σύμβολο εξόδου είναι ε . Αυτό συνεπάγεται ότι εάν δεν ορίζαμε τον πίνακα \mathbf{E} τότε θα χρειαζόταν να ελέγχουμε συνεχώς τους πίνακες \mathbf{I} και \mathbf{O} για να εξετάσουμε εάν και τα δύο σύμβολα είναι ε . Σε αυτή την περίπτωση, όμως, δεν θα μπορούσαμε να χρησιμοποιήσουμε την τροπική άλγεβρα για τον υπολογισμό, καθώς αυτός ο έλεγχος δεν μπορεί να γίνει 'αυτόματα' μέσω του πολλαπλασιασμού κάποιων πινάκων. Ο δεύτερος λόγος που καθιστά αναγκαίο τον ορισμό του πίνακα είναι ότι για τον ολικό υπολογισμό της ε -κλειστότητας χρειάζεται να υπολογίσουμε και τα βάρη κατά μήκος των μονοπατιών που περιέχουν μόνο ε -μεταβάσεις. Αυτό συνεπάγεται ότι πρέπει να αγνοήσουμε όλα τα άλλα βάρη, επειδή ο πίνακας $\Gamma(\mathbf{A})$ υπολογίζει τις ελάχιστες αποστάσεις μεταξύ των κόμβων, όμως οι αποστάσεις αυτές μπορεί να μην είναι κατά μήκος των μονοπατιών των ε -μεταβάσεων, αλλά να χρησιμοποιούν και άλλες ακμές.

Στο παράδειγμα του Σχήματος 4.4 βλέπουμε ότι το στοιχείο που αντιστοιχεί στην ελάχιστη απόσταση μεταξύ των κόμβων 0 και 2 του πίνακα $\Gamma(\mathbf{A})$ θα ήταν 0.30. Όμως, το βάρος



Σχήμα 4.4: Βοηθητική επεξήγηση της χρησιμότητας του πίνακα \mathbf{E} .

κατά μήκος του μονοπατιού που αποτελείται αποκλειστικά από ε -μεταβάσεις και ενώνει τους κόμβους 0 και 2 είναι 0.35 (και περνάει από τον κόμβο 1). Στο προκείμενο παράδειγμα, το αντίστοιχο στοιχείο του πίνακα $\Gamma(\mathbf{E})$ μας έδινε την αναμενόμενη τιμή.

Σημείωση 1. Στον υπολογισμό του πίνακα $\Gamma(\mathbf{A})$ για την ε -κλειστότητα θεωρούμε ως βάρος του μονοπατιού το ελάχιστο βάρος μεταξύ των εκάστοτε δύο κόμβων. Αυτή η επιλογή πηγάζει από τον περιορισμό των πολλαπλών μεταβάσεων: από τη στιγμή που δεν μπορούμε να έχουμε πολλαπλές μεταβάσεις από μια κατάσταση σε μια άλλη, κρατάμε απλώς αυτή με το μικρότερο κόστος. Τέτοιες περιπτώσεις θα μπορούσαν, εν δυνάμει, να είναι πληθώρες, στην περίπτωση που μια κατάσταση v έχει ε -μετάβαση απευθείας προς μια κατάσταση u , αλλά και μέσω μιας άλλης κατάστασης t .

Έχοντας υπολογίσει, πλέον, την ε -κλειστότητα του δικτύου μπορούμε να προχωρήσουμε στην ανανέωση των μεγεθών του WFST. Ας ξεκινήσουμε την ανάλυση με την εξέταση του νέου συνόλου των τελικών καταστάσεων. Αυτό, σύμφωνα με τον Αλγόριθμο 2, αποτελείται από τις αρχικές τελικές καταστάσεις, προσαυξημένο από τις καταστάσεις οι οποίες μπορούν να καταλήξουν σε κάποια τελική κατάσταση μέσω ε -μεταβάσεων. Ας θεωρήσουμε το διάνυσμα βαρών εκπομπής ρ , το οποίο έχει την τιμή των βαρών εκπομπής για κάθε μια από τις τελικές καταστάσεις, και το ουδέτερο στοιχείο ∞ εάν η αντίστοιχη κατάσταση δεν είναι τελική. Έστω τώρα το διάνυσμα ρ_ε :

$$\rho_\varepsilon = \Gamma(\mathbf{E}) \boxplus \rho \quad (4.6)$$

Ισχυρισμός 4.4. Το διάνυσμα ρ_ε της Εξίσωσης (4.6) περιέχει τα βάρη εκπομπής όλων των καταστάσεων οι οποίες γίνονται τελικές μόνο με ε -μεταβάσεις.

Ο ισχυρισμός επαληθεύεται από το γεγονός ότι ο πίνακας $\Gamma(\mathbf{E})$ εκφράζει όλα τα ε -μονοπάτια. Έστω ότι μια κατάσταση p μπορεί να μεταβεί σε μια κατάσταση q με ένα ε -μονοπάτι. Εάν, παράλληλα, η q αποτελεί τελική κατάσταση, τότε ο τροπικός πολλαπλασιασμός μεταξύ του πίνακα $\Gamma(\mathbf{E})$ και του διανύσματος ρ θα αντιστοιχίσει πεπερασμένη τιμή στο στοιχείο της κατάστασης p στο διάνυσμα ρ_ε (ίσο με το βάρος εκπομπής της κατάστασης q προσαυξημένο με το βάρος του ε -μονοπατιού από την κατάσταση p στην q), υποδηλώνοντας ότι και η p γίνεται τελική κατάσταση.

Έχοντας υπολογίσει τις καταστάσεις που γίνονται τελικές με ε -μεταβάσεις, μπορούμε να υπολογίσουμε το νέο διάνυσμα βαρών εκπομπής ως τον συνδυασμό του αρχικό διανύσματος ρ με το διάνυσμα ρ_ε :

$$\rho' = \rho_\varepsilon \wedge \rho = \Gamma(\mathbf{E}) \boxplus \rho \wedge \rho = \Gamma(\mathbf{E}) \boxplus \rho \wedge I_n \boxplus \rho = (\Gamma(\mathbf{E}) \wedge I_n) \boxplus \rho = \Delta(\mathbf{E}) \boxplus \rho \quad (4.7)$$

όπου I_n υποδηλώνει τον $n \times n$ ταυτοτικό πίνακα (δηλαδή έναν πίνακα με ταυτοτικά στοιχεία 0 στην διαγώνιο και ουδέτερα στοιχεία ∞ σε οποιοδήποτε άλλο δείκτη), και τα μη μηδενικά στοιχεία αυτού του τελικού διανύσματος υποδεικνύουν μια τελική κατάσταση.

Ας εξετάσουμε τώρα πως μεταβάλλεται ο πίνακας βαρών μετάβασης από την διαδικασία της αφαίρεσης ε -μεταβάσεων. Θεωρούμε τον πίνακα \mathbf{A}_ε :

$$(A_\varepsilon)_{ij} = \begin{cases} \infty, & \text{αν } I_{ij} = O_{ij} = \varepsilon \\ a_{ij}, & \text{διαφορετικά} \end{cases} \quad (4.8)$$

που αποτελεί, ουσιαστικά, τον συμπληρωματικό του πίνακα \mathbf{E} .

Ισχυρισμός 4.5. Ο πίνακας

$$\mathbf{A}' = \mathbf{A}_\varepsilon \wedge (\Gamma(\mathbf{E}) \boxplus \mathbf{A}_\varepsilon) = (\Gamma(\mathbf{E}) \wedge I_n) \boxplus \mathbf{A}_\varepsilon = \Delta(\mathbf{E}) \boxplus \mathbf{A}_\varepsilon \quad (4.9)$$

αποτελεί τον ανανεωμένο πίνακα βαρών μετάβασης.

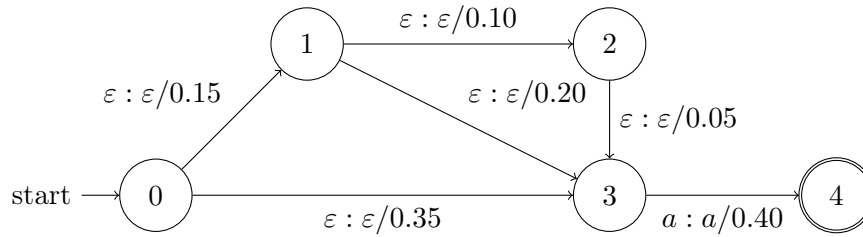
Ο ισχυρισμός επαληθεύεται από το γεγονός ότι ο πίνακας \mathbf{A}_ε αποτελείται τον πίνακα βαρών μετάβασης χωρίς τις ε -μεταβάσεις. Επαυξάνουμε αυτόν τον πίνακα με το γινόμενο $\Gamma(\mathbf{E}) \boxplus \mathbf{A}_\varepsilon$. Αυτό το βήμα είναι αναγκαίο επειδή ο πίνακας της ε -κλειστότητας $\Gamma(\mathbf{E})$ περιέχει μονάχα τα βάρη των ε -μεταβάσεων. Σε πλήρη αντιστοιχία με την επεξήγηση που δώσαμε όταν εισάγαμε τον Αλγόριθμο 2, αυτό το γεγονός συνεπάγεται ότι ο πίνακας $\Gamma(\mathbf{E})$ εμπεριέχει μόνο το βάρος κατά μήκος των μονοπατιών των ε -μεταβάσεων από την τωρινή κατάσταση σε μια επόμενη, χωρίς όμως να συμπεριλαμβάνει τη μετάβαση στην μετέπειτα κατάσταση. Ο σκοπός, όμως, του αλγορίθμου είναι να αντικαταστήσει πλήρως τις ε -μεταβάσεις με κανονικές μεταβάσεις, και όχι απλά να μεταβαίνει σε ισοδύναμες καταστάσεις. Αναφέρουμε τον αναγνώστη ξανά στο Σχήμα 4.3 για μια οπτική διευκρίνιση. Συνεπώς, ο πίνακας \mathbf{A}' αποτελεί τον ανανεωμένο πίνακα βαρών μετάβασης, καθώς εμπεριέχει τις αρχικές μεταβάσεις (που δεν αποτελούσαν ε -μεταβάσεις), καθώς και τις καινούργιες ακμές που εισήγαμε για να αντικαταστήσουμε τις ε -μεταβάσεις.

Παράδειγμα 4.2. Σε αυτή την ενότητα θα εξετάσουμε ένα παράδειγμα για να υπογραμμίσουμε τις διαφορές μεταξύ του Αλγορίθμου 2 για την αφαίρεση ε -μεταβάσεων, και των προτεινόμενων μοντελοποιήσεων των Εξισώσεων (4.7) και (4.9). Ας θεωρήσουμε το στοιχειώδες δίκτυο μιας τελικής κατάστασης του Σχήματος 4.5. Ο πίνακας βαρών μετάβασης του δικτύου είναι:

$$\mathbf{A} = \begin{bmatrix} \infty & 0.15 & \infty & 0.35 & \infty \\ \infty & \infty & 0.10 & 0.20 & \infty \\ \infty & \infty & \infty & 0.05 & \infty \\ \infty & \infty & \infty & \infty & 0.40 \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix}$$

Επιπλέον, αφού η μοναδική τελική κατάσταση είναι η κατάσταση 4, θα θεωρήσουμε ότι το διάνυσμα βαρών εκπομπής έχει την εξής μορφή:

$$\boldsymbol{\rho} = \begin{bmatrix} \infty \\ \infty \\ \infty \\ \infty \\ 5 \end{bmatrix}$$



Σχήμα 4.5: Δίκτυο για το παράδειγμα του αλγορίθμου αφαίρεσης ϵ -μεταβάσεων.

Ως ένα πρώτο βήμα θα εκτελέσουμε αναλυτικά τα βήματα του Αλγορίθμου 2. Διευκρινίζεται ότι με το σύμβολο \cup θα συμβολίζουμε την αθροιστική ένωση συνόλων. Αρχικοποιούμε το σύνολο των καινούργιων ακμών να είναι κενό, και έχουμε:

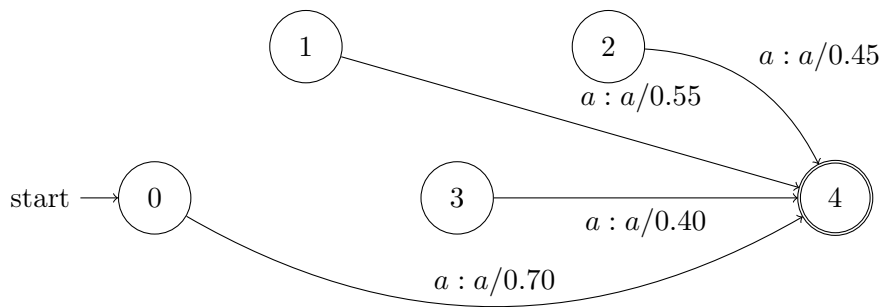
- Η εξεταζόμενη κατάσταση είναι η κατάσταση 0:
 $E' \leftarrow E' \cup \emptyset$, καθώς δεν υπάρχουν μη ϵ -μεταβάσεις από την κατάσταση 0. Η ϵ -κλειστότητα της εν λόγω κατάστασης είναι $EpsilonClosure(0) = \{1, 2, 3\}$:
 - $q = 1$ και $E' \leftarrow E' \cup \emptyset$ αφού η κατάσταση 1 έχει μόνο ϵ -μεταβάσεις.
 - $q = 2$ και $E' \leftarrow E' \cup \emptyset$ αφού η κατάσταση 2 έχει μόνο ϵ -μεταβάσεις.
 - $q = 3$ και $E' \leftarrow E' \cup (0, \alpha, \alpha, 0.30 + 0.4, 4)$ αφού το κόστος του συντομότερου ϵ -μονοπατιού μεταξύ των δύο καταστάσεων είναι 0.3 και η κατάσταση 3 μεταβαίνει στην κατάσταση 4 με σύμβολο εισόδου α , σύμβολο εξόδου α , και βάρος 0.4.
- Η εξεταζόμενη κατάσταση είναι η κατάσταση 1:
 $E' \leftarrow E' \cup \emptyset$, καθώς δεν υπάρχουν μη ϵ -μεταβάσεις από την κατάσταση 1. Η ϵ -κλειστότητα της εν λόγω κατάστασης είναι $EpsilonClosure(1) = \{2, 3\}$:
 - $q = 2$ και $E' \leftarrow E' \cup \emptyset$ αφού η κατάσταση 2 έχει μόνο ϵ -μεταβάσεις.
 - $q = 3$ και $E' \leftarrow E' \cup (1, \alpha, \alpha, 0.15 + 0.4, 4)$ αφού το κόστος του συντομότερου ϵ -μονοπατιού μεταξύ των δύο καταστάσεων είναι 0.15 και η κατάσταση 3 μεταβαίνει στην κατάσταση 4 με σύμβολο εισόδου α , σύμβολο εξόδου α , και βάρος 0.4.
- Η εξεταζόμενη κατάσταση είναι η κατάσταση 2:
 $E' \leftarrow E' \cup \emptyset$, καθώς δεν υπάρχουν μη ϵ -μεταβάσεις από την κατάσταση 2. Η ϵ -κλειστότητα της εν λόγω κατάστασης είναι $EpsilonClosure(2) = \{3\}$:
 - $q = 3$ και $E' \leftarrow E' \cup (2, \alpha, \alpha, 0.05 + 0.4, 4)$ αφού το κόστος του συντομότερου ϵ -μονοπατιού μεταξύ των δύο καταστάσεων είναι 0.05 και η κατάσταση 3 μεταβαίνει στην κατάσταση 4 με σύμβολο εισόδου α , σύμβολο εξόδου α , και βάρος 0.4.
- Η εξεταζόμενη κατάσταση είναι η κατάσταση 3:
 $E' \leftarrow E' \cup (3, \alpha, \alpha, 0.4, 4)$. Η ϵ -κλειστότητα της εν λόγω κατάστασης είναι κενή.
- Η εξεταζόμενη κατάσταση είναι η κατάσταση 4:
 $E' \leftarrow E' \cup \emptyset$ καθώς δεν υπάρχουν μεταβάσεις από την κατάσταση 4. Η ϵ -κλειστότητα της εν λόγω κατάστασης είναι κενή. Καθώς η κατάσταση είναι τελική, προστίθεται στο νέο σύνολο τελικών καταστάσεων.

Τελικά, το σύνολο ακμών που επιστρέφει ο Αλγόριθμος 2 είναι:

$$E' = \{(0, \alpha, \alpha, 0.7, 4), \\ (1, \alpha, \alpha, 0.55, 4), \\ (2, \alpha, \alpha, 0.45, 4), \\ (3, \alpha, \alpha, 0.4, 4)\}$$

Καθώς δεν υπάρχουν ε -μεταβάσεις προς την τελική κατάσταση 4, το νέο διάνυσμα βαρών εκπομπής ρ' ταυτίζεται με το αρχικό ρ .

Σε αυτό το σημείο παρατηρούμε κάτι ενδιαφέρον. Περιμέναμε ότι ο Αλγόριθμος 2 για την αφαίρεση ε -μεταβάσεων θα αγνοούσε πλήρως τις καταστάσεις οι οποίες είναι προσβάσιμες μονάχα ε -μεταβάσεις. Όμως, όπως φαίνεται στο Σχήμα 4.6 κάτι τέτοιο δεν συμβαίνει, και αντίθετα αυτές οι καταστάσεις (και οι ακμές τους) επιβιώνουν στο τελικό δίκτυο.

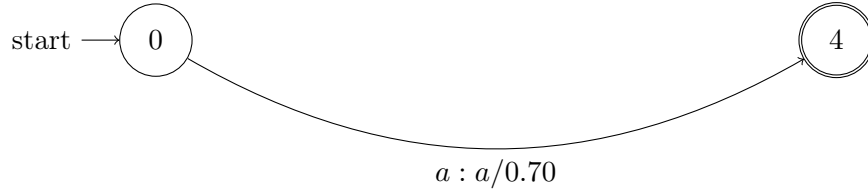


Σχήμα 4.6: Προκύπτων δίκτυο μετά την εφαρμογή του αλγορίθμου αφαίρεσης ε -μεταβάσεων.

Όμως ο αναγνώστης δεν χρειάζεται να θορυβείται, καθώς κάτι τέτοιο δεν αποτελεί κάποιο ουσιαστικό πρόβλημα. Η ιδιότητα αυτή των αλγορίθμων αφαίρεσης ε -μεταβάσεων είναι γνωστή και δεν είναι δυνατό να αντιμετωπιστεί κατά τη διάρκεια της εκτέλεσης του αλγορίθμου, καθώς, εάν μεταβάλλαμε τη δομή του WFST κατά την εκτέλεση του βρόχου της γραμμής 1, τότε θα οδηγούμασταν σε λάθος αποτέλεσμα. Οι συγγραφείς του [29] ξεπερνούν αυτό το εμπόδιο εφαρμόζοντας έναν αλγόριθμο ψαλιδίσματος στο τελικό, χωρίς ε -μεταβάσεις WFST, το οποίο αφαιρεί 'νεκρές' καταστάσεις και μεταβάσεις (όπου με τον όρο 'νεκρές' εννοούμε καταστάσεις οι οποίες δεν είναι προσβάσιμες από κάποια αρχική κατάσταση). Θα αποφύγουμε να εξετάσουμε επιμελώς έναν τέτοιο αλγόριθμο. Αλγόριθμοι στους οποίους μεταβάλλεται ο αριθμός των καταστάσεων είναι αρκετά δύσκολο να περιγραφούν μέσω δομών δεδομένων στατικών διαστάσεων, όπως είναι οι πίνακες. Συνεπώς, η περιγραφή δεν θα ενέπιπτε στον επιθυμητό φορμαλισμό, και συνεπώς είναι ξεπερνάει το εύρος ενδιαφέροντος.

Παρά τα παραπάνω, θα κάνουμε μια σύντομη νύξη στο πως, πρακτικά, θα μπορούσαμε γρήγορα να υπολογίσουμε κάτι τέτοιο στο πλαίσιο της τροπικής άλγεβρας. Εάν θεωρήσουμε τον προκύπτων πίνακα μεταβάσεων \mathbf{A}' (αφού έχει ολοκληρωθεί κάποιος αλγόριθμος αφαίρεσης ε -μεταβάσεων) τότε ο πίνακας $\Gamma(\mathbf{A}')$ εκφράζει τις ελάχιστες αποστάσεις μεταξύ όλων των κόμβων του δικτύου. Έστω $I = \{q_1, q_2, \dots, q_\lambda\}$ το σύνολο όλων των αρχικών καταστάσεων. Εάν όλες οι γραμμές που αντιστοιχούν στις καταστάσεις του συνόλου I έχουν το στοιχείο ∞ στη στήλη που αντιστοιχεί σε μια κατάσταση p στον πίνακα $\Gamma(\mathbf{A}')$, αυτό συνεπάγεται ότι η p δεν είναι προσβάσιμη από καμία αρχική κατάσταση, και άρα μπορεί να αφαιρεθεί.

Στο παράδειγμά του Σχήματος 4.6, η εφαρμογή ενός αλγορίθμου ψαλιδίσματος θα είχε ως αποτέλεσμα την εξάλειψη των καταστάσεων 1, 2, και 3 (καθώς αυτές οι καταστάσεις δεν είναι



Σχήμα 4.7: Προκύπτων δίκτυο μετά την εφαρμογή του αλγορίθμου ψαλιδίσματος.

προσβάσιμες με κανέναν τρόπο από την αρχική κατάσταση), και ο τελικός γράφος παρουσιάζεται στο Σχήμα 4.7.

Θα επαναλάβουμε τώρα το προηγούμενο παράδειγμα σύμφωνα με τις Εξισώσεις (4.7) και (4.9). Αρχικά θα υπολογίσουμε τους πίνακες \mathbf{A}' και \mathbf{E} της Εξίσωσης (4.4). Έχουμε:

$$\mathbf{A} = \begin{bmatrix} \infty & 0.15 & \infty & 0.35 & \infty \\ \infty & \infty & 0.10 & 0.20 & \infty \\ \infty & \infty & \infty & 0.05 & \infty \\ \infty & \infty & \infty & \infty & 0.40 \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix}, \quad \mathbf{A}_\varepsilon = \begin{bmatrix} \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & 0.40 \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix},$$

$$\mathbf{E} = \begin{bmatrix} \infty & 0.15 & \infty & 0.35 & \infty \\ \infty & \infty & 0.10 & 0.20 & \infty \\ \infty & \infty & \infty & 0.05 & \infty \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix}$$

Μπορούμε πλέον να υπολογίσουμε την ε -κλειστότητα όλων των καταστάσεων του WFST σύμφωνα με την Εξίσωση (4.5):

$$\Gamma(\mathbf{E}) = \begin{bmatrix} \infty & 0.15 & 0.25 & 0.30 & \infty \\ \infty & \infty & 0.10 & 0.15 & \infty \\ \infty & \infty & \infty & 0.05 & \infty \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix}$$

Έχοντας υπολογίσει όλους τους απαραίτητους πίνακες, μπορούμε να προχωρήσουμε στον υπολογισμό των νέων μεγεθών του WFST, ρ' και \mathbf{A}' . Ας υπολογίσουμε, αρχικά, το ενδιάμεσο γινόμενο μεταξύ των πινάκων $\Gamma(\mathbf{E})$ και \mathbf{A}_ε , έτσι ώστε να υπάρχει μεγαλύτερη διαφάνεια στους υπολογισμούς:

$$\Gamma(\mathbf{E}) \boxplus \mathbf{A}_\varepsilon = \begin{bmatrix} \infty & 0.15 & 0.25 & 0.30 & \infty \\ \infty & \infty & 0.10 & 0.15 & \infty \\ \infty & \infty & \infty & 0.05 & \infty \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix} \boxplus \begin{bmatrix} \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & 0.40 \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix} =$$

$$= \begin{bmatrix} \infty & \infty & \infty & \infty & 0.70 \\ \infty & \infty & \infty & \infty & 0.55 \\ \infty & \infty & \infty & \infty & 0.45 \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix}$$

Τελικά, ο νέος πίνακας μεταβάσεων \mathbf{A}' της Εξίσωσης (4.9) παίρνει την μορφή:

$$\begin{aligned} \mathbf{A}' = \mathbf{A}_\varepsilon \wedge (\Gamma(\mathbf{E}) \boxplus \mathbf{A}_\varepsilon) &= \begin{bmatrix} \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & 0.40 \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix} \wedge \begin{bmatrix} \infty & \infty & \infty & \infty & 0.70 \\ \infty & \infty & \infty & \infty & 0.55 \\ \infty & \infty & \infty & \infty & 0.45 \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix} = \\ &= \begin{bmatrix} \infty & \infty & \infty & \infty & 0.70 \\ \infty & \infty & \infty & \infty & 0.55 \\ \infty & \infty & \infty & \infty & 0.45 \\ \infty & \infty & \infty & \infty & 0.40 \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix} \end{aligned}$$

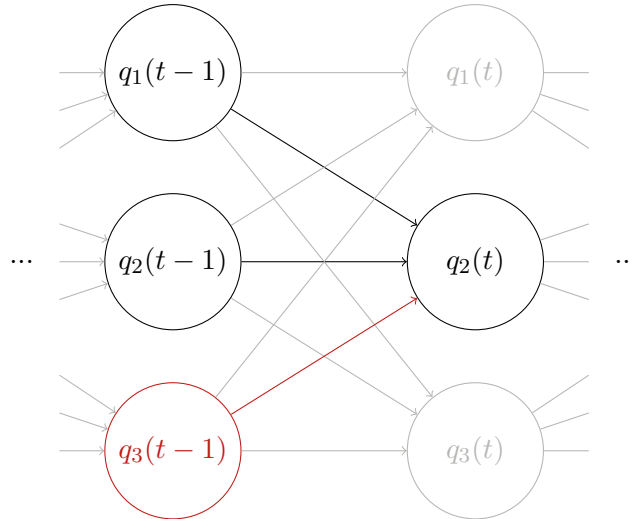
Το νέο διάνυσμα βαρών εκπομπής $\boldsymbol{\rho}'$ προκύπτει από την Εξίσωση (4.7):

$$\boldsymbol{\rho}' = \Gamma(\mathbf{E}) \boxplus \boldsymbol{\rho} \wedge \boldsymbol{\rho} = \begin{bmatrix} \infty & 0.15 & 0.25 & 0.30 & \infty \\ \infty & \infty & 0.10 & 0.15 & \infty \\ \infty & \infty & \infty & 0.05 & \infty \\ \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix} \boxplus \begin{bmatrix} \infty \\ \infty \\ \infty \\ \infty \\ 5 \end{bmatrix} \wedge \begin{bmatrix} \infty \\ \infty \\ \infty \\ \infty \\ 5 \end{bmatrix} = \begin{bmatrix} \infty \\ \infty \\ \infty \\ \infty \\ 5 \end{bmatrix}$$

Όπως ήταν αναμενόμενο, το WFST που υπονοείται από τον πίνακα μεταβάσεων \mathbf{A}' και το διάνυσμα βαρών εκπομπής $\boldsymbol{\rho}'$ ταυτίζεται με το δίκτυο του Σχήματος 4.6. Και πάλι μπορούμε να εκτιμήσουμε την απλότητα των υπολογισμών, καθώς και την κομψότητα του φορμαλισμού των Εξισώσεων (4.7) και (4.9).

4.3 Αλγόριθμος Viterbi

Σε αυτή την ενότητα προχωράμε στην εξέταση του, ίσως, σημαντικότερου αλγορίθμου που εφαρμόζεται στα γραφικά μοντέλα, τον αλγόριθμο Viterbi ([20], [47], [52]). Όπως ήδη αναφέραμε στην εισαγωγή του κεφαλαίου, ο αλγόριθμος Viterbi χρησιμοποιείται για την εύρεση των πιο πιθανών μονοπατιών σε ένα γραφικό μοντέλο. Απλούστερα, δεδομένου μιας ακολουθίας διανυσμάτων παρατηρήσεων, παράγει την ακολουθία καταστάσεων του μοντέλου η οποία 'ταιριάζει' καλύτερα στα δεδομένα. Αντίθετα με τους αλγορίθμους που έχουμε εξετάσει μέχρι στιγμής στα Χωρία 4.1 και 4.2, οι οποίοι μεταβάλλουν τη δομή του WFST, ο εν λόγω αλγόριθμος αποτελεί έναν μη επεμβατικό αλγόριθμο ο οποίος απλώς διατρέχει το δίκτυο και το χρησιμοποιεί κατά την εκτέλεσή του, αφήνοντας παντελώς αναλλοίωτη τη δομή του. Σε κάθε βήμα ο αλγόριθμος υπολογίζει για κάθε κατάσταση την τιμή μέγιστης πιθανότητας (ή ελαχίστου βάρους στον φορμαλισμό της τροπικής άλγεβρας) με βάση τα βάρη μετάβασης και τις τιμές των μεγίστων πιθανοτήτων του προηγούμενου βήματος. Έχοντας εντοπίσει το βέλτιστο μονοπάτι, ανανεώνει την τιμή σύμφωνα με τις πιθανότητες παρατήρησης, και αποθηκεύει το αποτέλεσμα ως την μέγιστη πιθανότητα του τρέχοντος βήματος. Στο Σχήμα 4.8 παρουσιάζεται μια οπτική αναπαράσταση της παραπάνω διαδικασίας. Συγκεκριμένα, στο παράδειγμα του σχήματος υποθέτουμε ότι προσπαθούμε να υπολογίσουμε την τιμή που θα αναθέσουμε στο $q_2(t)$. Γι' αυτόν τον υπολογισμό εξετάζουμε όλες τις τιμές του προηγούμενου βήματος (συνυπολογίζοντας τα βάρη μετάβασης), και επιλέγουμε την μέγιστη (εδώ υποθέτουμε ότι αυτή προέρχεται από την διαφορετικού χρώματος κατάσταση $q_3(t-1)$).



Σχήμα 4.8: Συμβολικό παράδειγμα εφαρμογής του αλγορίθμου Viterbi.

4.3.1 Ανάλυση

Ο κλασικός αλγόριθμος Viterbi μπορεί να γραφτεί στην εξής max-product μορφή:

$$q_i(t) = \left(\max_j w_{ji} q_j(t-1) \right) \cdot b_i(\sigma_t) \quad (4.10)$$

όπου w_{ji} είναι η πιθανότητα μετάβασης από την κατάσταση j στην κατάσταση i , $b_i(\sigma_t)$ είναι η πιθανότητα να παρατηρηθεί το σύμβολο εισόδου σ_t (στο χρονικό πλαίσιο t), ενώ βρισκόμαστε στην κατάσταση i , και $q_i(t)$ είναι το καλύτερο score (μέγιστη πιθανότητα) κατά μήκος ενός μονοπατιού το οποίο, βελτιστοποιημένο κατά τα προηγούμενα $t-1$ χρονικά πλαίσια, καταλήγει στην κατάσταση i κατά το χρονικό πλαίσιο t και ενσαρκώνει τα παρατηρούμενα σύμβολα εισόδου $\sigma_1, \dots, \sigma_t$.

Εν αντιθέσει με τα προηγούμενα χωρία όπου αναλύσαμε εκτενώς την λειτουργία των αλγορίθμων πριν προχωρήσουμε στην μοντελοποίησή τους στο πεδίο της τροπικής άλγεβρας, στο προκείμενο θα προχωρήσουμε κατευθείαν στην μοντελοποίηση. Ο λόγος γι' αυτό είναι ότι, αφ' ενός, ο αλγόριθμος Viterbi είναι από τους πιο διαδεδομένους αλγορίθμους της Επιστήμης των Υπολογιστών, και συνεπώς δεν χρειάζεται τις ίδιες εκτενείς εισαγωγές. Αφ' εταίρου, η λειτουργία του είναι πιο εύκολα κατανοητή, καθώς αποσκοπεί στην εύρεση βέλτιστων μονοπατιών.

Προχωρώντας στην μοντελοποίηση, θα χρειαστεί να πάρουμε τον αρνητικό λογάριθμο της (4.10), έτσι ώστε να μεταβούμε από το πεδίο της max-product άλγεβρας στο πεδίο της τροπικής άλγεβρας. Έχουμε:

$$-\log q_i(t) = \left(\min_j (-\log w_{ji} q_j(t-1)) \right) - \log b_i(\sigma_t) \quad (4.11)$$

Τώρα, θέτοντας $\mathbf{x}(t) = -\log \mathbf{q}(t)$, $\mathbf{A} = -\log \mathbf{W}$, και $\mathbf{p}(\sigma_t) = -\log \mathbf{b}(\sigma_t)$ μπορούμε να γράψουμε την Εξίσωση (4.11) σε συμβολισμό πινάκων:

$$\mathbf{x}(t) = \mathbf{A}^T \boxplus \mathbf{x}(t-1) + \mathbf{p}(\sigma_t) \quad (4.12)$$

Παρατηρούμε, όμως, ότι μεταξύ των διανυσμάτων $\mathbf{A}^T \boxplus \mathbf{x}(t-1)$ και $\mathbf{p}(\sigma_t)$ εμφανίζεται η πράξη της συμβατικής πρόσθεσης. Όπως αναλύσαμε εκτενώς στο Κεφάλαιο 2, μια τέτοια

πράξη δεν είναι καλώς ορισμένη μεταξύ διανυσμάτων (και πινάκων) στο πεδίο της τροπικής άλγεβρας. Σε όλη την μέχρι τώρα ανάλυσή μας θέλαμε να είμαστε αυστηροί ως προς τη χρήση μόνο πράξεων της τροπικής άλγεβρας. Προς αυτή την κατεύθυνση, έστω ο πίνακας $\mathbf{P}(\sigma_t)$ τον οποίον ορίζουμε ως εξής:

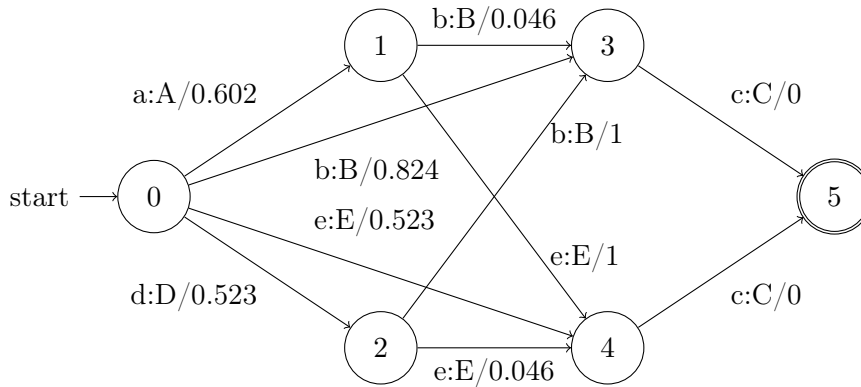
$$\mathbf{P}(\sigma_t) = \begin{bmatrix} p_1(\sigma_t) & \infty & \cdots & \infty \\ \infty & p_2(\sigma_t) & \cdots & \infty \\ \vdots & \vdots & \ddots & \vdots \\ \infty & \infty & \cdots & p_n(\sigma_t) \end{bmatrix} \quad (4.13)$$

Πλέον, χρησιμοποιώντας τον πίνακα $\mathbf{P}(\sigma_t)$ της Εξίσωσης (4.13) η Εξίσωση (4.12) μπορεί να γραφεί ως εξής:

$$\mathbf{x}(t) = \mathbf{P}(\sigma_t) \boxplus \mathbf{A}^T \boxplus \mathbf{x}(t-1) \quad (4.14)$$

Ο φορμαλισμός της Εξίσωσης (4.14) είναι, αποκλειστικά πλέον, στο πεδίο της τροπικής άλγεβρας.

Παράδειγμα 4.3. Ας εξετάσουμε ένα παράδειγμα για να κατανοήσουμε περαιτέρω την φορμαλισμό της Εξίσωσης (4.14) και να υπογραμμίσουμε τις υπολογιστικές διαφορές από την Εξίσωση (4.10). Έστω το δίκτυο του Σχήματος 4.9. Ο πίνακας μεταβάσεων \mathbf{A} που υπονοε-



Σχήμα 4.9: Δίκτυο για το παράδειγμα εφαρμογής του αλγορίθμου Viterbi.

ίται από τα βάρη των ακμών του σχήματος είναι:

$$\mathbf{A} = \begin{bmatrix} \infty & 0.602 & 0.523 & 0.824 & 0.523 & \infty \\ \infty & \infty & \infty & 0.046 & 1 & \infty \\ \infty & \infty & \infty & 1 & 0.046 & \infty \\ \infty & \infty & \infty & \infty & \infty & 0 \\ \infty & \infty & \infty & \infty & \infty & 0 \\ \infty & \infty & \infty & \infty & \infty & \infty \end{bmatrix}$$

Επιπλέον, θα υποθέσουμε ότι κάθε κατάσταση έχει υψηλότερη πιθανότητα να παρατηρήσει το σύμβολο εισόδου της (το σύμβολο, δηλαδή, που υποδεικνύεται στο Σχήμα 4.9) από το να παρατηρήσει τα υπόλοιπα σύμβολα. Ποσοτικοποιώντας την προηγούμενη δήλωση, έστω ότι η πιθανότητα του συμβόλου εισόδου είναι 0.30 για κάθε σύμβολο στην αντίστοιχη κατάσταση, και 0.175 για τα υπόλοιπα σύμβολα. Τότε η κατάσταση 1 έχει υψηλή πιθανότητα να παρατηρήσει το σύμβολο εισόδου a , η κατάσταση 2 έχει υψηλή πιθανότητα να παρατηρήσει το

σύμβολο b , και ούτω καθ' εξής. Αναφορικά με την κατάσταση 0, υποθέτουμε ότι δεν είναι δυνατό να παρατηρήσει κανένα σύμβολο (και συνεπώς η πιθανότητα παρατήρησης οποιουδήποτε συμβόλου είναι 0). Παίρνοντας τον αρνητικό λογάριθμο αυτών των πιθανοτήτων, τα βάρη, πλέον, παρατήρησης γίνονται 0.523 για το σύμβολο εισόδου και 0.757 για τα άλλα σύμβολα σε κάθε κατάσταση. Για τον 'κλασικό' υπολογισμό του αλγορίθμου Viterbi θα χρησιμοποιήσουμε την εξής εξίσωση:

$$x_i(t) = \min_j (a_{ji} + x_j(t-1)) + p_i(\sigma_t) \quad (4.15)$$

Ο λόγος που επιλέγουμε να χρησιμοποιήσουμε την Εξίσωση (4.15) (έναντι της Εξίσωσης (4.10)) είναι ότι έτσι τα αποτελέσματα θα είναι άμεσα συγκρίσιμα με τα επόμενα, όπου για τους υπολογισμούς θα εφαρμόσουμε την Εξίσωση (4.14). Επιπλέον, θα γράφουμε τους υπολογισμούς σε μορφή διανυσμάτων για λόγους ευκρίνειας, οι υπολογισμοί, όμως, θα γίνονται πάντα σύμφωνα με την Εξίσωση (4.15). Το αρχικό διάνυσμα καταστάσεων, αφού η κατάσταση 0 είναι η μόνη δυνατή αρχική κατάσταση, είναι:

$$\mathbf{x}(0) = \begin{bmatrix} 0 \\ \infty \\ \infty \\ \infty \\ \infty \\ \infty \end{bmatrix}$$

Εστω ότι παρατηρούμε τα σύμβολα a , b , και c , στα πρώτα τρία χρονικά πλαίσια. Το διάνυσμα καταστάσεων μετά την πρώτη επανάληψη του αλγορίθμου παίρνει τη μορφή:

$$\mathbf{x}(1) = \begin{bmatrix} \min_j (a_{j0} + x_j(0)) + p_0(a) \\ \min_j (a_{j1} + x_j(0)) + p_1(a) \\ \min_j (a_{j2} + x_j(0)) + p_2(a) \\ \min_j (a_{j3} + x_j(0)) + p_3(a) \\ \min_j (a_{j4} + x_j(0)) + p_4(a) \\ \min_j (a_{j5} + x_j(0)) + p_5(a) \end{bmatrix} = \begin{bmatrix} (\infty + 0) + \infty \\ (0.602 + 0) + 0.523 \\ (0.523 + 0) + 0.757 \\ (0.824 + 0) + 0.757 \\ (0.523 + 0) + 0.757 \\ (\infty + 0) + 0.757 \end{bmatrix} = \begin{bmatrix} \infty \\ 1.125 \\ 1.28 \\ 1.581 \\ 1.28 \\ \infty \end{bmatrix}$$

Προχωρώντας στην δεύτερη επανάληψη, το διάνυσμα καταστάσεων είναι:

$$\mathbf{x}(2) = \begin{bmatrix} \min_j (a_{j0} + x_j(1)) + p_0(b) \\ \min_j (a_{j1} + x_j(1)) + p_1(b) \\ \min_j (a_{j2} + x_j(1)) + p_2(b) \\ \min_j (a_{j3} + x_j(1)) + p_3(b) \\ \min_j (a_{j4} + x_j(1)) + p_4(b) \\ \min_j (a_{j5} + x_j(1)) + p_5(b) \end{bmatrix} = \begin{bmatrix} (\infty + 1.125) + \infty \\ (\infty + 1.125) + 0.757 \\ (\infty + 1.125) + 0.757 \\ (0.046 + 1.125) + 0.523 \\ (0.046 + 1.28) + 0.757 \\ (0 + 1.28) + 0.757 \end{bmatrix} = \begin{bmatrix} \infty \\ \infty \\ \infty \\ 1.694 \\ 2.083 \\ 2.037 \end{bmatrix}$$

Τέλος, μετά το πέρας της τρίτης επανάληψης το διάνυσμα καταστάσεων έχει την εξής μορφή:

$$\mathbf{x}(3) = \begin{bmatrix} \min_j (a_{j0} + x_j(2)) + p_0(c) \\ \min_j (a_{j1} + x_j(2)) + p_1(c) \\ \min_j (a_{j2} + x_j(2)) + p_2(c) \\ \min_j (a_{j3} + x_j(2)) + p_3(c) \\ \min_j (a_{j4} + x_j(2)) + p_4(c) \\ \min_j (a_{j5} + x_j(2)) + p_5(c) \end{bmatrix} = \begin{bmatrix} (\infty + 1.694) + \infty \\ (\infty + 1.694) + 0.757 \\ (\infty + 1.694) + 0.757 \\ (\infty + 1.694) + 0.757 \\ (\infty + 1.694) + 0.757 \\ (0 + 1.694) + 0.523 \end{bmatrix} = \begin{bmatrix} \infty \\ \infty \\ \infty \\ \infty \\ \infty \\ 2.217 \end{bmatrix}$$

Έχοντας ολοκληρώσει την εκτέλεση του αλγορίθμου Viterbi σύμφωνα με την Εξίσωση (4.15), προχωράμε στην επανάληψη του παραδείγματος με τη χρήση του φορμαλισμού της

τροπικής άλγεβρας, μέσω της Εξίσωσης (4.14). Αρχικά, θα κατασκευάσουμε τον πίνακα $\mathbf{P}(\sigma_t)$ της Εξίσωσης (4.13) για κάθε χρονικό πλαίσιο. Οι πίνακες παρατηρήσεων των τριών χρονικών πλαισίων είναι:

$$\mathbf{P}(a) = \begin{bmatrix} \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & 0.523 & \infty & \infty & \infty & \infty \\ \infty & \infty & 0.757 & \infty & \infty & \infty \\ \infty & \infty & \infty & 0.757 & \infty & \infty \\ \infty & \infty & \infty & \infty & 0.757 & \infty \\ \infty & \infty & \infty & \infty & \infty & 0.757 \end{bmatrix},$$

$$\mathbf{P}(b) = \begin{bmatrix} \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & 0.757 & \infty & \infty & \infty & \infty \\ \infty & \infty & 0.757 & \infty & \infty & \infty \\ \infty & \infty & \infty & 0.523 & \infty & \infty \\ \infty & \infty & \infty & \infty & 0.757 & \infty \\ \infty & \infty & \infty & \infty & \infty & 0.757 \end{bmatrix},$$

$$\mathbf{P}(c) = \begin{bmatrix} \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & 0.757 & \infty & \infty & \infty & \infty \\ \infty & \infty & 0.757 & \infty & \infty & \infty \\ \infty & \infty & \infty & 0.757 & \infty & \infty \\ \infty & \infty & \infty & \infty & 0.757 & \infty \\ \infty & \infty & \infty & \infty & \infty & 0.523 \end{bmatrix}$$

Σε αυτό το σημείο κρίνουμε σκόπιμο, όπως κάναμε και στα προηγούμενα χωρία, να επαναλάβουμε τον πίνακα βαρών μετάβασης \mathbf{A} , καθώς και το αρχικό διάνυσμα καταστάσεων $\mathbf{x}(0)$, ώστε να διευκολυνθεί ο αναγνώστης στην παρακολούθηση του παραδείγματος:

$$\mathbf{A} = \begin{bmatrix} \infty & 0.602 & 0.523 & 0.824 & 0.523 & \infty \\ \infty & \infty & \infty & 0.046 & 1 & \infty \\ \infty & \infty & \infty & 1 & 0.046 & \infty \\ \infty & \infty & \infty & \infty & \infty & 0 \\ \infty & \infty & \infty & \infty & \infty & 0 \\ \infty & \infty & \infty & \infty & \infty & \infty \end{bmatrix}, \quad \mathbf{x}(0) = \begin{bmatrix} 0 \\ \infty \\ \infty \\ \infty \\ \infty \\ \infty \end{bmatrix}$$

Στο βωμό του χώρου και της κομψότητας, δεν θα εκτελέσουμε αναλυτικά τους πολλαπλασιασμούς πινάκων γράφοντας αναλυτικά τους επιμέρους πίνακες, όπως στα προηγούμενα χωρία, καθώς, εν προκειμένω, συμμετέχουν στους πολλαπλασιασμούς περισσότεροι πίνακες. Επομένως, το διάνυσμα καταστάσεων μετά την πρώτη επανάληψη παίρνει τη μορφή:

$$\mathbf{x}(1) = \mathbf{P}(a) \boxplus \mathbf{A}^T \boxplus \mathbf{x}(0) = \begin{bmatrix} \infty \\ 1.125 \\ 1.28 \\ 1.581 \\ 1.28 \\ \infty \end{bmatrix}$$

Αντίστοιχα, μετά το πέρας της δεύτερης επανάληψης, το διάνυσμα καταστάσεων διαμορφώνεται

ως εξής:

$$\mathbf{x}(2) = \mathbf{P}(b) \boxplus \mathbf{A}^T \boxplus \mathbf{x}(1) = \begin{bmatrix} \infty \\ \infty \\ \infty \\ 1.694 \\ 2.083 \\ 2.037 \end{bmatrix}$$

Τέλος, με την ολοκλήρωση της τρίτης επανάληψης, το διάνυσμα καταστάσεων είναι:

$$\mathbf{x}(2) = \mathbf{P}(b) \boxplus \mathbf{A}^T \boxplus \mathbf{x}(1) = \begin{bmatrix} \infty \\ \infty \\ \infty \\ \infty \\ \infty \\ 2.217 \end{bmatrix}$$

Παρατηρούμε ότι, ακόμη μια φορά, τα αποτελέσματα των δύο διαφορετικών προσεγγίσεων ταυτίζεται. Εκτιμούμε ότι ο φορμαλισμός της Εξίσωσης (4.14) διευκολύνει τους χειρωνακτικούς υπολογισμούς, καθώς "αυτοματοποιεί" την διαδικασία συνδυασμού των μεγεθών και του υπολογισμού των ελαχίστων. Συγκεκριμένα, χρησιμοποιώντας την Εξίσωση (4.15) χρειάζεται να συνδυάσουμε χειροκίνητα τα επιμέρους a_{ij} και $x_j(t-1)$, και έπειτα να υπολογίσουμε το ελάχιστο αυτών, κάτι το οποίο στον φορμαλισμό της Εξίσωσης (4.14) προκύπτει οργανικά από τον τροπικό πολλαπλασιασμό πινάκων.

4.4 Αλγόριθμος Viterbi με αποκοπή

Ο αλγορίθμος Viterbi με αποκοπή αποτελεί μια παραλλαγή του κλασσικού αλγορίθμου Viterbi. Σε αυτή την παραλλαγή, σε κάθε βήμα του αλγορίθμου, αποκόβονται κάποια μονοπάτια (δηλαδή οι τιμές των καταστάσεων τίθενται στο ουδέτερο στοιχείο του εκάστοτε 'πολλαπλασιασμού') και δεν εξετάζονται σε περαιτέρω εκτελέσεις. Ο λόγος που επιλέγεται η χρήση του εν λόγω αλγορίθμου είναι διττός:

- Αφ' ενός, είναι σύνηθες στα συστήματα όπου εφαρμόζεται ο αλγόριθμος Viterbi να υπάρχουν απαιτήσεις για εκτέλεση πραγματικού χρόνου. Καθώς τα δίκτυα σε αυτές τις εφαρμογές είναι αρκετά μεγάλα, και υπάρχουν οι χρονικοί περιορισμοί, η αποκοπή μονοπατιών με κάποια, πιθανώς, επίδραση στην απόδοση κρίνεται προτιμότερη από τον αργότερο χρόνο εκτέλεσης.
- Λόγω της δομής των εν λόγω δικτύων, συχνά κάποια μονοπάτια (ή κάποιες συστάδες μονοπατιών) είναι εντελώς απίθανα δεδομένων των παρατηρήσεων. Όμως, στον κλασικό αλγόριθμο Viterbi αυτά τα μονοπάτια θα συνεχίσουν να εξετάζονται με την ελπίδα ότι θα βελτιωθεί η πιθανότητά τους σε περαιτέρω βήματα, κάτι το οποίο συνήθως δεν συμβαίνει, και έτσι απλώς εισάγουν επιπλέον υπολογιστικό φόρτο στον αλγόριθμο.

Η εφαρμογή της παραλλαγής του αλγορίθμου Viterbi έχει ως αποτέλεσμα την άμεση βελτίωση του χρόνου εκτέλεσης της αποκωδικοποίησης. Εκτός, όμως, της βελτίωσης του χρόνου εκτέλεσης μέσω της αγνόησης μερικών μονοπατιών, η χρήση του αλγορίθμου επιτρέπει την εκμετάλλευση του αλγορίθμου προώθησης βαρών που εισήχθη στην Ενότητα 4.1, οδηγώντας σε ακόμα σημαντικότερη βελτίωση του χρόνου εκτέλεσης.

Προχωρώντας στο πρακτικό κομμάτι, η αποκοπή γίνεται συνήθως με ένα από δύο κριτήρια, ή ακόμα και με τον συνδυασμό τους:

- προσδιορίζεται από το χρήστη μια παράμετρος ‘χαλάρωσης’ θ . Σε κάθε βήμα του αλγορίθμου υπολογίζεται η κατάσταση k με το χαμηλότερο βάρος (δηλαδή η πιο πιθανή κατάσταση) και οι μόνες καταστάσεις που επιβιώνουν τη διαδικασία της αποκοπής είναι αυτές οι οποίες απέχουν το πολύ θ από την πιθανότερη κατάσταση k .
- προσδιορίζεται από τον χρήστη έναν αριθμός n . Σε κάθε βήμα του αλγορίθμου ταξινομούνται οι καταστάσεις σε αύξουσα σειρά ως προς το βάρος τους, και οι μόνες καταστάσεις που επιβιώνουν τη διαδικασία της αποκοπής είναι οι n καταστάσεις με το χαμηλότερο βάρος.

Όταν χρησιμοποιείται ο συνδυασμός τους, συνήθως επιβιώνουν οι καταστάσεις οι οποίες απέχουν το πολύ θ από τη βέλτιστη, και εισάγεται ο περιορισμός ο αριθμός αυτών των καταστάσεων που θα επιβιώσουν να είναι το πολύ n .

4.4.1 Ανάλυση

Υπενθυμίζουμε στον αναγνώστη την μοντελοποίηση του αλγορίθμου Viterbi στην οποία είχαμε καταλήξει στην Εξίσωση (4.14):

$$\mathbf{x}(t) = \mathbf{P}(\sigma_t) \boxplus \mathbf{A}^T \boxplus \mathbf{x}(t-1)$$

όπου \mathbf{A} είναι ο πίνακας βαρών μετάβασης, $\mathbf{P}(\sigma_t)$ ο πίνακας βαρών παρατήρησης, και $\mathbf{x}(t-1)$ το διάνυσμα καταστάσεων του προηγούμενου πλαισίου.

Στην παραλλαγή του αλγορίθμου Viterbi με αποκοπή, μετά το πέρας των υπολογισμών της Εξίσωσης (4.14) για κάθε χρονικό πλαίσιο t , διατρέχουμε σειριακά το διάνυσμα $\mathbf{x}(t)$ και θέτουμε τα στοιχεία τα οποία έχουν τιμή μεγαλύτερη από ένα κατώφλι, το οποίο εξαρτάται από την τιμή της παραμέτρου χαλάρωσης θ και την ελάχιστη τιμή του διανύσματος $\mathbf{x}(t)$, σε $+\infty$ (το ουδέτερο στοιχείο του πολλαπλασιασμού της τροπικής άλγεβρας). Υποστηρίζουμε ότι οι δείκτες των καταστάσεων οι οποίοι θα πρέπει να αποβιώσουν κατά τη διαδικασία της αποκοπής μπορούν να προσδιοριστούν μέσω της χρήσης του αντίστροφου Cuninghame-Green ([15], [40]).

Πρόταση 4.1. Έστω

$$\mathbf{X}(t) = \begin{bmatrix} x_1(t) & \infty & \cdots & \infty \\ \infty & x_2(t) & \cdots & \infty \\ \vdots & \vdots & \ddots & \vdots \\ \infty & \infty & \cdots & x_n(t) \end{bmatrix}$$

όπου $x_i(t)$ αντιπροσωπεύει το i -οστό στοιχείο του διανύσματος $\mathbf{x}(t)$, και έστω $\boldsymbol{\eta} = \theta + \frac{1}{2} (\mathbf{x}(t)^T \boxplus \mathbf{x}(t)) + \mathbf{0}$, όπου $\mathbf{0}$ είναι ένα διάνυσμα του οποίου όλα τα στοιχεία είναι 0 και θ είναι η παράμετρος χαλάρωσης. Τέλος, έστω ότι το σύμβολο \boxplus' εκφράζει τον *max-plus* πολλαπλασιασμό πινάκων και $\mathbf{X}^\#(t) := -\mathbf{X}^T(t)$ (αντίστροφος Cuninghame-Green). Τότε, τα αρνητικά στοιχεία της ποσότητας

$$\bar{\mathbf{y}} = \mathbf{X}^\#(t) \boxplus' \boldsymbol{\eta} \quad (4.16)$$

υποδεικνύουν τους δείκτες του διανύσματος $\mathbf{x}(t)$ οι οποίοι χρειάζεται να αποκοπούν.

Απόδειξη. Η λύση $\bar{\mathbf{y}}$ που προκύπτει από την χρήση του αντίστροφου Cuninghame-Green παρέχει την μικρότερη δυνατή λύση στην ανισότητα $\mathbf{X}(t) \boxplus \mathbf{y} \geq \boldsymbol{\eta}$. Όμως, ανάλογα με τους

διαγώνιους πίνακες της γραμμικής άλγεβρας, ο διαγώνιος πίνακας $\mathbf{X}(t)$ αντιστοιχίζει κάθε ένα από τα στοιχεία του $x_i(t)$ στο εκάστοτε στοιχείο y_i . Το διάνυσμα που προκύπτει αποτελείται από αθροίσματα ζευγών y_i και $x_i(t)$, τα οποία χρειάζεται να είναι μεγαλύτερα από την τιμή του κατωφλίου. Όμως, όπως αναφέραμε το διάνυσμα $\bar{\mathbf{y}}$ αποτελεί την μικρότερη δυνατή λύση, και συνεπώς εάν το κάποιο $x_i(t)$ είναι ήδη μεγαλύτερο από την τιμή του κατωφλίου, τότε το \bar{y}_i θα δεχθεί αρνητική τιμή, υποδηλώνοντας ότι ο δείκτης ο οποίος αντιστοιχεί στο $x_i(t)$ ξεπερνάει το κατώφλι, και κατά συνέπεια πρέπει να αποκοπεί. \square

Στον ορισμό του $\boldsymbol{\eta}$ στην Εξίσωση (4.16) εισήγαμε την ποσότητα $\frac{1}{2} (\mathbf{x}(t)^T \boxplus \mathbf{x}(t))$. Το γινόμενο $\mathbf{x}(t)^T \boxplus \mathbf{x}(t)$ αποτελεί ουσιαστικά την τροπική νόρμα του διανύσματος $\mathbf{x}(t)$, και επιστρέφει το διπλάσιο του ελαχίστου στοιχείου του διανύσματος (εξ' ου και η εισαγωγή του πολλαπλασιαστικού παράγοντα $\frac{1}{2}$ στον υπολογισμό). Η εισήγηση της ποσότητας $\mathbf{0}$ γίνεται καθώς όλα τα υπόλοιπα μεγέθη της ποσότητας είναι βαθμωτά, και συνεπώς είναι απαραίτητη έτσι ώστε το $\boldsymbol{\eta}$ να είναι, εν τέλει, διάνυσμα. Ας δούμε ένα σύντομο παράδειγμα. Έστω το διάνυσμα:

$$\mathbf{x} = \begin{bmatrix} 1 \\ 7 \\ 4 \end{bmatrix}$$

Ας υποθέσουμε, επιπλέον, ότι η παράμετρος χαλάρωσης είναι $\theta = 5$. Τότε το διάνυσμα $\boldsymbol{\eta}$ της Εξίσωσης (4.16) παίρνει τη μορφή:

$$\boldsymbol{\eta} = 5 + \frac{1}{2} (2 \times \min(1, 7, 4)) + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \\ 6 \end{bmatrix}$$

Η βέλτιστη λύση της ανισότητας $\mathbf{X} \boxplus \mathbf{y} \geq \boldsymbol{\eta}$ (όπου, όπως εξηγήσαμε, $\mathbf{X} = \text{diag}(\mathbf{x})$) δίνεται από την Εξίσωση (4.16):

$$\bar{\mathbf{y}} = \begin{bmatrix} -1 & -\infty & -\infty \\ -\infty & -7 & -\infty \\ -\infty & -\infty & -4 \end{bmatrix} \boxplus' \begin{bmatrix} 6 \\ 6 \\ 6 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix}$$

Σύμφωνα με την Πρόταση 4.1, αφού το στοιχείο \bar{y}_2 είναι αρνητικό, τότε δεν πρέπει να επιβιώσει την διαδικασία της αποκοπής. Έτσι, το διάνυσμα που θα αποδώσει τελικά ο αλγόριθμος Viterbi με αποκοπή είναι:

$$\mathbf{x}' = \begin{bmatrix} 1 \\ \infty \\ 4 \end{bmatrix}$$

4.4.2 Γεωμετρία

Σε αντίθεση με τα προηγούμενα χωρία, στο προκείμενο θα αφιερώσουμε χώρο στην μελέτη της γεωμετρίας του αλγορίθμου Viterbi με αποκοπή. Συγκεκριμένα, αποσκοπούμε στην εξαγωγή κάποιας γεωμετρικής δομής από τις εξισώσεις του Viterbi, την οποία επιθυμούμε να εκμεταλλευτούμε μετέπειτα.

Οδεύοντας προς αυτόν τον στόχο, ας θεωρήσουμε ένα διάνυσμα μεταβλητών \mathbf{z} . Ο λόγος για αυτήν τον ενέργεια είναι διττός:

- Αφ' ενός, η χρήση ενός διανύσματος μεταβλητών μας απελευθερώνει από την άμεση μεταβολή του εκάστοτε διανύσματος $\mathbf{x}(t)$. Κάτι τέτοιο είναι επιθυμητό, καθώς τα στοιχεία του $\mathbf{x}(t)$ καθορίζονται μονοσήμαντα από την Εξίσωση (4.14), και μετέπειτα μεταβολή αυτών δημιουργεί σύγχυση.

- Αφ' εταίρου οι γεωμετρικοί χώροι της αλγεβρικής γεωμετρίας (και, κατ' επέκταση, της τροπικής γεωμετρίας) ορίζονται ως λύσεις συνόλων εξισώσεων, γεγονός που καθιστά επιτακτική την εισαγωγή μεταβλητών, ώστε να ορίζεται κάποιος χώρος λύσεων στις εξισώσεις που επιδρούν πάνω στις μεταβλητές.

Θα φράξουμε τις πιθανές τιμές του διανύσματος μεταβλητών \mathbf{z} χρησιμοποιώντας τον νόμο ανανέωσης του Viterbi. Για να το επιτύχουμε αυτό, θα θεωρήσουμε την εξής χαλάρωση της Εξίσωσης (4.10):

$$\mathbf{z} \geq \mathbf{b}, \quad \mathbf{b} = \mathbf{P}(\sigma_t) \boxplus \mathbf{A}^T \boxplus \mathbf{x}(t-1) \quad (4.17)$$

Ο λόγος που επιλέξαμε την χαλάρωση της μορφής \geq έναντι της μορφής \leq είναι εξαιτίας της σημασίας των δύο μορφών ανισώσεων στο πλαίσιο της τροπικής άλγεβρας. Η ανισότητα $\mathbf{G} \boxplus \mathbf{z} \leq \mathbf{b}$ μεταφράζεται ως:

‘Το ελάχιστο στοιχείο της i -στης γραμμής του γινομένου μεταξύ των \mathbf{G} και \mathbf{z} χρειάζεται να είναι **μικρότερο** από b_i .’

Όμως, αυτή η ανισότητα δεν παρέχει ουσιώδη περιορισμούς στο διάνυσμα μεταβλητών \mathbf{z} . Αναλύοντας την επιρροή της, καταλαβαίνουμε ότι επηρεάζει μονάχα ένα στοιχείο του διανύσματος, και δεν είναι δυνατό να γνωρίζουμε, στη γενική περίπτωση, πιο στοιχείο επηρεάζεται. Εναλλακτικά, ας θεωρήσουμε την ανισότητα $\mathbf{G} \boxplus \mathbf{z} \geq \mathbf{b}$. Αυτή η ανισότητα μεταφράζεται ως:

‘Το ελάχιστο στοιχείο της i -στης γραμμής του γινομένου μεταξύ των \mathbf{G} και \mathbf{z} χρειάζεται να είναι **μεγαλύτερο** από b_i .’

Αφού το ελάχιστο στοιχείο ενός διανύσματος είναι μεγαλύτερο από μια τιμή, έπεται ότι κάθε στοιχείο του γινομένου είναι μεγαλύτερο από αυτή την τιμή.

Πλέον έχουμε φράξει το διάνυσμα μεταβλητών \mathbf{z} χρησιμοποιώντας τον νόμο ανανέωσης του Viterbi. Ο απώτερος σκοπός μας είναι να ορίσουμε ένα γεωμετρικό σύνολο. Για λόγους συμβατότητας με τον μαθηματικό ορισμό που δώσαμε στο Κεφάλαιο 2 για τα πολύτοπα, θα θέσουμε $\mathbf{G} = \mathbf{I}_n$. Έτσι, η Εξίσωση (4.17) μπορεί να γραφεί στη μορφή:

$$\mathbf{G} \boxplus \mathbf{z} \geq \mathbf{b} \quad (4.18)$$

Στην Εξίσωση (4.16) αντιμετωπίσαμε την αποκοπή ως ένα πρόβλημα εκτίμησης, όπου προσπαθήσαμε να εκτιμήσουμε τους δείκτες προς αποκοπή μέσω του αντίστροφου Cuninghame-Green. Ο λόγος γι' αυτό ήταν ότι δεν είχαμε κάποιον τρόπο να επιβάλουμε τον νόμο ανανέωσης του Viterbi της Εξίσωσης (4.10) στο διάνυσμα \mathbf{y} . Όμως, το διάνυσμα μεταβλητών \mathbf{z} έχει περιοριστεί από την χαλάρωση της Εξίσωσης (4.17), και άρα η διαδικασία της αποκοπής μπορεί να εξεταστεί υπό ένα νέο πρίσμα. Μπορούμε να θεωρήσουμε τη διαδικασία της αποκοπής ως άλλο ένα σύνολο περιορισμών στο διάνυσμα μεταβλητών \mathbf{z} . Πράγματι:

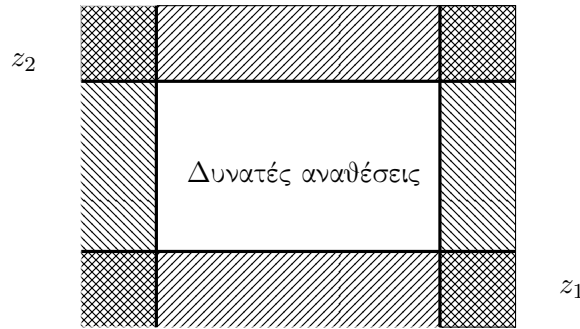
$$\mathbf{z} \leq \boldsymbol{\eta}, \quad \boldsymbol{\eta} = \boldsymbol{\eta} + \frac{1}{2} (\mathbf{b}^T \boxplus \mathbf{b}) + \mathbf{0} \quad (4.19)$$

όπου το διάνυσμα \mathbf{b} ορίστηκε στην Εξίσωση (4.17). Ακολουθώντας το ίδιο τέχνασμα που επιστρατεύσαμε στην Εξίσωση (4.18), έστω $\mathbf{H} = \mathbf{I}_n$. Τότε η (4.19) μπορεί να γραφεί στην παρακάτω μορφή:

$$\mathbf{H} \boxplus \mathbf{z} \leq \boldsymbol{\eta} \quad (4.20)$$

Πλέον έχουμε στη διάθεσή μας όλα τα απαραίτητα εργαλεία ώστε να ορίσουμε ένα γεωμετρικό σύνολο λύσεων. Ο συνδυασμός των Εξισώσεων (4.18) και (4.20) ορίζει ένα πολύτοπο, συμβατό με τον ορισμό του Κεφαλαίου 2. Πράγματι:

$$T(\mathbf{G}^b, \mathbf{H}^\eta) = \{\mathbf{z} \in \mathbb{R}_{\min}^n : \mathbf{G}^b \boxplus \mathbf{z} \geq \mathbf{0}, \quad \mathbf{H}^\eta \boxplus \mathbf{z} \leq \mathbf{0}\} \quad (4.21)$$



Σχήμα 4.10: Δυνατές αναθέσεις περιορισμένες από τον νόμο ανανέωσης του Viterbi (κατώτεροι και αριστερότεροι περιορισμοί) και από την διαδικασία αποκοπής (άνωτεροι και δεξιότεροι περιορισμοί).

όπου ο πίνακας \mathbf{G}^b είναι πανομοιότυπος με τον πίνακα \mathbf{G} , με την μοναδική εξαίρεση ότι τα στοιχεία κάθε γραμμής i έχουν μειωθεί κατά b_i , και ο πίνακας \mathbf{H}^n είναι πανομοιότυπος με τον πίνακα \mathbf{H} , με την μοναδική εξαίρεση ότι τα στοιχεία κάθε γραμμής i έχουν μειωθεί κατά n_i .

Ο χώρος που εγκλείεται από το πολύτοπο της Εξίσωσης (4.21) περιέχει όλες τις δυνατές αναθέσεις που μπορεί να πάρει το διάνυσμα μεταβλητών \mathbf{z} οι οποίες ικανοποιούν τους περιορισμούς. Ένα διαισθητικό παράδειγμα του περιορισμένου χώρου οπτικοποιείται στο Σχήμα 4.10.

4.4.3 Ερμηνεία

Έχοντας ορίσει τον γεωμετρικό χώρο λύσεων του αλγορίθμου Viterbi με αποκοπή, αφιερώνουμε αυτή την ενότητα στην εξέταση της ερμηνείας των γεωμετρικών ιδιοτήτων του χώρου, καθώς και την εξαγωγή πληροφοριών αναφορικά με την αποτελεσματικότητα της διαδικασίας αποκοπής από το πολύτοπο και την δομή του.

Ας θεωρήσουμε τον εγκλειόμενο 'όγκο' στο εσωτερικό του πολυτόπου (αναφερόμαστε σε d -διάστατο όγκο, δηλαδή ένα ευθύγραμμο τμήμα στον χώρο 1- Δ , μια επιφάνεια στον χώρο 2- Δ , τον συμβατικό όγκο στον χώρο 3- Δ , κτλ). Ο όγκος παρέχει πληροφορία αναφορικά με το μέγεθος του χώρου λύσεων. Υποστηρίζουμε ότι αυτός ο όγκος μπορεί να δώσει μια αίσθηση της αποτελεσματικότητας της διαδικασίας αποκοπής. Όμως, ο όγκος εξαρτάται άμεσα από τον συνολικό αριθμό των συμμετεχόντων κόμβων. Συνεπώς θα ήταν συνετό να ενσωματώσουμε στις μετρικές το υποστήριγμα του αποκομμένου διανύσματος μεταβλητών \mathbf{z} .

Ορισμός 4.4. Το υποστήριγμα ενός διανύσματος \mathbf{x} , που συμβολίζεται με $\text{supp}(\mathbf{x})$ είναι το σύνολο των δεικτών που αντιστοιχούν σε πεπερασμένα στοιχεία του διανύσματος \mathbf{x} .

Επιπλέον, ας θεωρήσουμε την ποσότητα $r_i = (\min(\mathbf{z}) + \eta) - z_i, \forall i$. Τότε καταλαβαίνουμε ότι $r_i \in [0, \eta]$, και ουσιαστικά το r_i αποτελεί έναν δείκτη του κατά πόσο κάθε δείκτης του διανύσματος \mathbf{z} ικανοποιεί τους περιορισμούς των Εξισώσεων (4.18) και (4.20). Εάν, δηλαδή, κάποιο r_i είναι μεγάλο, αυτό συνεπάγεται ότι η τιμή της αντίστοιχης διάστασης ικανοποιεί εύκολα τους περιορισμούς των εξισώσεων. Αντίθετα, εάν η τιμή κάποιου r_i είναι ιδιαίτερα μικρή, τότε η εν λόγω διάσταση οριακά δεν αποκόβεται κατά τη διαδικασία της αποκοπής. Τότε:

- Έστω ν ο αρνητικός λογάριθμος του κανονικοποιημένου όγκου, διαιρεμένος με το υπο-

στήριγμα του αποκομμένου διανύσματος:

$$\nu = -\frac{1}{\text{supp}(\mathbf{z})} \log \left(\prod_{i \in \text{supp}(\mathbf{z})} \frac{r_i}{\max \mathbf{r}} \right) \quad (4.22)$$

$$= -\frac{1}{\text{supp}(\mathbf{z})} \sum_{i \in \text{supp}(\mathbf{z})} \frac{\log r_i}{\log (\max \mathbf{r})} \quad (4.23)$$

- Έστω ε :

$$\varepsilon = -\frac{1}{\text{supp}(\mathbf{z})} \sum_{i \in \text{supp}(\mathbf{z})} -z_i(t) \cdot e^{-z_i(t)} \quad (4.24)$$

Ουσιαστικά, αυτή η μετρική είναι η εντροπία του διανύσματος πιθανοτήτων $\mathbf{q}(t)$. Θυμίζουμε ότι $\mathbf{x} = -\log \mathbf{q}$, έχουμε:

$$-\sum_{i \in \text{supp}(\mathbf{z})} -z_i(t) \cdot e^{-z_i(t)} = -\sum_{i \in \text{supp}(\mathbf{z})} q_i(t) \cdot \log q_i(t)$$

Θα προβούμε τώρα στην συνοπτική ανάλυση των παραπάνω μετρικών. Συγκεκριμένα, θα προσπαθήσουμε να εξηγήσουμε πως ακριβώς διαμορφώνονται τα μεγέθη και τι ακριβώς συνεπάγονται οι τιμές που επιδέχονται. Συγκεκριμένα:

- Έστω ο κανονικοποιημένος όγκος ν της Εξίσωσης (4.22). Εάν ο κανονικοποιημένος όγκος που εσωκλείεται στο εσωτερικό του προκύπτον πολυτόπου είναι μεγάλος, αυτό συνεπάγεται ότι τα επιμέρους κλάσματα επιδέχονται μικρές τιμές (και άρα η εφαρμογή του αρνητικού λογαρίθμου οδηγεί μεγάλες τιμές, συνεπάγοντας μεγάλο όγκο). Τότε, ο χώρος πιθανών βέλτιστων λύσεων έχει μια ιδιαίτερη μορφή: υπάρχουν αρκετές διαστάσεις στο διάνυσμα μεταβλητών \mathbf{z} οι οποίες δεν είναι ιδιαίτερα πιθανές, καθώς τους έχουν ανατεθεί σημαντικά μεγαλύτερες τιμές αρνητικών λογαρίθμων, συγκριτικά με τι βέλτιστη. Επομένως, σε τέτοια περίπτωση, η αποκοπή έχει περιθώρια να γίνει πιο ‘επιθετική’, δηλαδή δύναται να μεγαλώσουμε την τιμή που αναθέτουμε στην παράμετρο χαλάρωσης θ , αποκόπτοντας, έτσι, περισσότερα πιθανά μονοπάτια. Αντίθετα, εάν ο κανονικοποιημένος όγκος είναι μικρός, αυτό συνεπάγεται ότι τα επιμέρους κλάσματα έχουν τιμές κοντά στη μονάδα, και άρα ο αρνητικός λογάριθμος των κανονικοποιημένων κλασμάτων επιδέχεται μικρές τιμές. Τότε ο χώρος λύσεων αποτελείται από διαστάσεις που είναι, σε γενικές γραμμές, ισοπίθανες. Επομένως, σε τέτοια περίπτωση η αποκοπή είναι πιθανώς υπέρμετρα επιθετική, και υπάρχει σημαντικός κίνδυνος να αποκοπεί το βέλτιστο μονοπάτι. Συνεπώς, θα ήταν φρόνιμο σε τέτοια περίπτωση να αυξήσουμε την τιμή της παραμέτρου χαλάρωσης, έτσι ώστε να οδηγηθούμε στην επιβίωση περισσότερων πιθανών μονοπατιών.
- Έστω η κανονικοποιημένη εντροπία ε της Εξίσωσης (4.24). Η τιμή της εντροπίας έχει μεγάλες τιμές, τότε αυτό συνεπάγεται ότι η πληροφορία που δύναται να εξαχθεί για την εν λόγω κατανομή είναι μεγάλη. Ίδεατά, θα θέλαμε η συνολική εντροπία να τείνει στο μηδέν, καθώς κάτι τέτοιο θα υποδήλωνε ότι δεν μπορούμε να εξάγουμε επιπλέον πληροφορία από τα δείγματα που είναι διαθέσιμα, και άρα είμαστε σίγουροι για την βελτιστότητα των μονοπατιών. Όμως, κάτι τέτοιο δεν είναι εφικτό, για ποικιλία λόγων. Αρχικά, λόγω του θορύβου έμφυτου στα προβλήματα και της διακύμανσης των κατανομών, αλλά και των ανακριβειών της μοντελοποίησης, πάντα θα υπάρχουν δείγματα τα

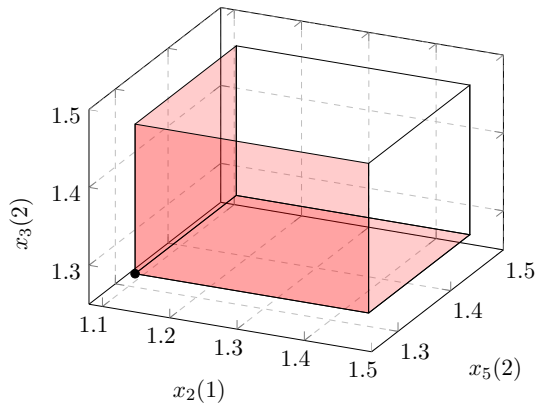
οποία οδηγούν σε μετρήσιμες τιμές εντροπίας. Παράλληλα, λόγω της φύσης των προβλημάτων που εξετάζουμε, όταν οι παρατηρήσεις αντιστοιχούν σε σπάνιες λέξεις, έναντι των πιο πιθανών λέξεων, ανεξάρτητα των παρατηρήσεων, οι υπολογισμοί θα οδηγήσουν σε παρατήρηση μεγάλων τιμών εντροπίας. Έτσι, πρακτικά, δεν δύναται να αποσκοπούμε στην ελαχιστοποίηση της εντροπίας. Όμως, μπορούμε να αποσκοπούμε στην ελαχιστοποίηση του ρυθμού μεταβολής της. Κάτι τέτοιο έχει ως αποτέλεσμα ότι εξασφαλίζουμε ότι τα νέα δείγματα τα οποία εξετάζουμε δεν οδηγούν σε σημαντική πληροφορία για την μορφή των κατανομών, και άρα βρισκόμαστε σε τοπικό ελάχιστο για την τιμή της παραμέτρου χαλάρωσης θ .

Με βάση την παραπάνω ανάλυση, μπορούμε να βγάλουμε μερικά γενικά συμπεράσματα για την χρήση των μετρικών. Εν γένει, θα θέλαμε να αποφύγουμε να έχουμε μικρό κανονικοποιημένο όγκο ν , καθώς, όπως αναφέραμε, κάτι τέτοιο ενέχει υπαρκτό κίνδυνο να αποκοπεί το βέλτιστο μονοπάτι. Παράλληλα, όμως, η ανεξέλεγκτη και αλόγιστη μεγιστοποίηση του κανονικοποιημένου όγκου οδηγεί σε απαγορευτικούς χρόνους εκτέλεσης, καθώς ολοένα αυξάνεται ο αριθμός των εξεταζόμενων μονοπατιών σε κάθε χρονικό πλαίσιο. Η χρήση της μετρικής της εντροπίας ε έρχεται να ‘φρενάρει’ την ανεξέλεγκτη αύξηση του κανονικοποιημένου όγκου. Συγκεκριμένα, στην Θεωρία Πληροφορίας ([14], [48], [50]) η εντροπία εκφράζει την αξιοποιήσιμη πληροφορία που μπορούμε να εξάγουμε από μια κατανομή πιθανότητας. Συγκεκριμένα, η εντροπία αποτελεί μια ποσοτικοποίηση της ‘έκπληξης’ από την παρατήρηση τιμών σε κάποιο πείραμα. Έτσι, εάν παρατηρούμε τιμές οι οποίες, με βάση την πρόωμη γνώση, συνάδουν πλήρως με τα στοιχεία της κατανομής, τότε αυτές οι τιμές θα οδηγούν σε μικρές τιμές εντροπίας. Ο λόγος γι’ αυτό είναι ότι, από τη στιγμή που αυτές δεν αντιτίθενται στην μοντελοποίηση που έχουμε για την κατανομή, δεν εκπλήσσουν, και άρα δεν προσφέρουν και επιπλέον πληροφορία, καθώς δεν αλλάζει η εκτίμησή μας για την κατανομή. Αντίθετα, η παρατήρηση ενός δείγματος το οποίο αντιθέτει ρητά τα πρώιμα δείγματα θα έχει πολύ μεγάλη τιμή εντροπίας, και θα παρέχει σημαντική πληροφορία, καθώς υποδηλώνει ότι η κατανομή, πιθανώς, να συμπεριφέρεται διαφορετικά από ότι περιμέναμε. Συνεπώς, θα θέλαμε, μέσω των αλγορίθμων που θα αναπτύξουμε και των επιλογών που θα κάνουμε για τις τιμές της παραμέτρου χαλάρωσης θ , να μεγιστοποιούμε την μετρική του κανονικοποιημένου όγκου, όσο, όμως, κάτι τέτοιο οδηγεί σημαντική διαφορά στην μετρική της εντροπίας. Έτσι, θα βελτιστοποιήσουμε την πληροφορία την οποία κερδίζουμε με την επιλογή να αυξήσουμε το χρόνο εκτέλεσης μέσω της εισαγωγής του νέου όγκου. Αντίθετα, εάν παρατηρούμε αύξηση του όγκου, χωρίς σημαντική διαφορά στις τιμές της μετρικής ε τότε καταλήγουμε ότι δεν αποκτήσαμε επιπλέον πληροφορία, και, κατά συνέπεια, η αύξηση της παραμέτρου χαλάρωσης ήταν αντιπαραγωγική.

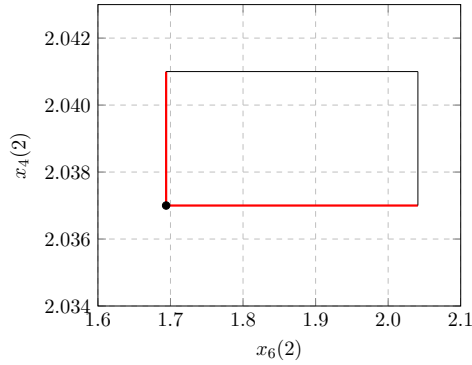
Αυτές οι μετρικές παρέχουν διαισθήσεις αναφορικά με την αποτελεσματικότητα της διαδικασίας αποκοπής, και μπορούν να βοηθήσουν τους σχεδιαστές συστημάτων να διαλέξουν τιμές για την παράμετρο χαλάρωσης θ . Τόσο η μετρική του κανονικοποιημένου όγκου ν όσο και η μετρική της εντροπίας ε μπορούν να χρησιμοποιηθούν για να ποσοτικοποιήσουμε την ισορροπία μεταξύ της υπολογιστικής πολυπλοκότητας του αλγορίθμου Viterbi με αποκοπή, και της πιθανότητας εύρεσης του βέλτιστου μονοπατιού. Η ελαχιστοποίηση του ρυθμού μεταβολής της μετρικής ε , παράλληλα με τη μεγιστοποίηση της μετρικής ν μπορούν να βοηθήσουν άμεσα στην επιλογή των τιμών των παραμέτρων χαλάρωσης, τόσο σε στατικούς σχεδιασμούς συστημάτων, όσο και στην ανάπτυξη αλγορίθμων για την αυτόματη προσαρμογή της παραμέτρου θ .

Ενδιαφέρον έχει το πως το υποστήριγμα του διανύσματος \mathbf{x} σχετίζεται με το υποστήριγμα του διανύσματος \mathbf{z} . Συγκεκριμένα:

- εξετάζοντας το μέγεθος της ποσότητας $\text{supp}(\mathbf{z})$ συγκριτικά με την ποσότητα $\text{supp}(\mathbf{x})$, μπορούμε να αλλάζουμε δυναμικά την τιμή της παραμέτρου χαλάρωσης θ μεταξύ των διαφορετικών πλαισίων.



(α) Το πολύτοπο του πρώτου πλαισίου.



(β') Το πολύτοπο του δεύτερου πλαισίου.

Σχήμα 4.11: Τα προκύπτον πολύτοπα του Παραδείγματος 4.4.

- μπορούμε να εξετάσουμε το σύνολο V των καταστάσεων για τις οποίες ισχύει ότι:

$$V = \{i : i \in \text{supp}(\mathbf{x}), i \notin \text{supp}(\mathbf{z})\}$$

Είναι δυνατό να βρούμε ομοιότητες σε αυτές τις καταστάσεις (και στα αποδεχόμενα μονοπάτια τους), και κατά συνέπεια να απλοποιήσουμε περαιτέρω τα WFSTs.

Μια τελευταία θεώρηση είναι να εξετάσουμε πώς οι κορυφές του πολυτόπου συνδέονται με τα N -καλύτερα μονοπάτια τα οποία συνήθως παράγονται από αλγορίθμους αποκοπής. Για την ακρίβεια, κάποιες $n - 1$ -πλευρές του πολυτόπου εμφανίζονται στα N -καλύτερα μονοπάτια. Αυτές οι $n - 1$ -πλευρές παράγονται από τον κατώτερο περιορισμό για κάθε διάσταση, και συνεπώς ενεργούν πάνω στον χώρο ο οποίο παράγεται από τις άλλες διαστάσεις. Οι κόμβοι όπου αυτές οι υπερεπιφάνειες τέμνονται είναι το διάνυσμα το οποίο επιστρέφει ο αλγόριθμος, και περιέχει όλα τα βέλτιστα μονοπάτια για το εκάστοτε πλαίσιο.

Παράδειγμα 4.4. Έστω ότι η παράμετρος χαλάρωσης είναι $\theta = 0.347$ (που ισούται με τον αρνητικό λογάριθμο του 0.45). Χρησιμοποιώντας την εξίσωση (4.19), τα διανύσματα κατωφλίωσης αποτελούνται από τα στοιχεία 1.472 και 2.041 αντίστοιχα. Τα πολύτοπα που ορίζονται από τον συνδυασμό των περιορισμών του νόμου ανανέωσης του Viterbi και τους περιορισμούς της παραλλαγής της αποκοπής για τα διανύσματα $\mathbf{x}(1)$ και $\mathbf{x}(2)$ μπορούν να βρεθούν στο Σχήμα 4.11.

Κεφάλαιο 5

Πειραματισμός

Έχοντας παρουσιάσει το απαραίτητο θεωρητικό υπόβαθρο στα Κεφάλαια 2 και 3, και έχοντας αναπτύξει τις κύριες ιδέες που διέπουν την εν λόγω διπλωματική εργασία στο Κεφάλαιο 4, θα αφιερώσουμε αυτό το κεφάλαιο σε πειραματική μελέτη. Συγκεκριμένα, θα παρουσιάσουμε πληθώρα πειραμάτων που αποσκοπούν να υπογραμμίσουν την πρακτική εφαρμογή των ιδεών που αναπτύξαμε στα προηγούμενα κεφάλαια, και ιδιαίτερα στην Ενότητα 4.4. Αρχικά, θα παρουσιάσουμε μερικά πειράματα που έγιναν σε ένα σύνολο δεδομένων για Επεξεργασία Φυσικής Γλώσσας χρησιμοποιώντας τις ιδέες των πολυτόπων της Ενότητας 4.4. Το πείραμα αποσκοπεί στην μείωση της χρονικής πολυπλοκότητας του αλγορίθμου Viterbi με αποκοπή, κάνοντας χρήση των μετρικών ε και ν έτσι ώστε να γίνει ορθή επιλογή της παραμέτρου χαλάρωσης θ . Έπειτα, θα προχωρήσουμε στον εκτενέστερο πειραματισμό μέσω προσομοιώσεων. Συγκεκριμένα, θα παρουσιάσουμε τρεις διακριτές οικογένειες πειραμάτων, οι οποίες αποσκοπούν στην ορθή επιλογή της παραμέτρου χαλάρωσης, αξιοποιώντας τις μετρικές του Κεφαλαίου 4, και στην αυτόματη προσαρμογή της παραμέτρου χαλάρωσης.

5.1 Πειράματα σε Επεξεργασία Φυσικής Γλώσσας

Το πεδίο της Επεξεργασίας Φυσικής Γλώσσας αποτελεί ένα πρόσφορο έδαφος για την εφαρμογή του αλγορίθμου Viterbi με αποκοπή. Οι λόγοι γι' αυτό μπορούν να αναλυθούν στους παρακάτω κύριους άξονες:

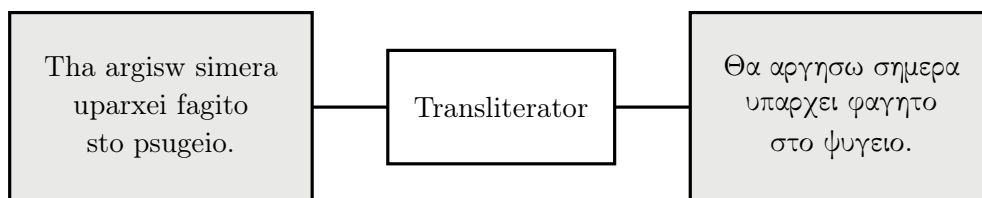
- Αρχικά, η παραμετροποίηση των μοντέλων γλώσσας δύναται να υλοποιηθεί οργανικά μέσω WFSTs. Έτσι, ο τομέας, στο σύνολό του, επιδέχεται την εφαρμογή του αλγορίθμου Viterbi χωρίς να απαιτείται σημαντική μετατροπή της μοντελοποίησης των δεδομένων.
- Επιπλέον, η μοντελοποίηση είναι αρκετά φυσική, καθώς σε πολλά προβλήματα της Επεξεργασίας Φυσικής Γλώσσας αναδύεται η ανάγκη για μετατροπή μιας δοσμένης ακολουθίας συμβόλων σε κάποια άλλη. Χαρακτηριστικό παράδειγμα αποτελούν τα προβλήματα της περίληψης, της μετάφρασης, κ.ά.

Με αυτό το έναυσμα, πειραματιστήκαμε με την υλοποίηση των αλγορίθμων του Κεφαλαίου 4, και την μετέπειτα εφαρμογή τους σε ένα σύνολο δεδομένων ενός πρόβλημα του τομέα της Επεξεργασίας Φυσικής Γλώσσας. Συγκεκριμένα, εφαρμόσαμε τις ιδέες της Ενότητας 4.4 αναφορικά με τον ορισμό πολυτόπων στα διάφορα χρονικά πλαίσια που προκύπτουν κατά την εκτέλεση του αλγορίθμου Viterbi με αποκοπή, καθώς και της χρήσης των μετρικών της κανονικοποιημένης εντροπίας ε και του κανονικοποιημένου όγκου ν σε ένα πρόβλημα transliteration ([36], [18]). Στα προβλήματα του είδους το ζητούμενο είναι η μετατροπή ενός

κειμένου μιας συγκεκριμένης γλώσσας γραμμένο με ένα αλφάβητο α σε κείμενο της ίδιας γλώσσας, γραμμένο με ένα αλφάβητο β . Τέτοια προβλήματα προκύπτουν πολύ συχνά στη σύγχρονη εποχή, ειδικά σε γλώσσες οι οποίες δεν χρησιμοποιούν (επακριβώς) το λατινικό αλφάβητο, και συνήθως πηγάζει από την επιθυμία για γρήγορη και αβίαστη επικοινωνία μεταξύ των χρηστών. Θα αναφέρουμε δύο χαρακτηριστικά παραδείγματα στα οποία προκύπτει οργανικά η ανάγκη του transliteration. Τα δύο παραδείγματα διαφέρουν ως προς τον άξονα ότι στο πρώτο η χρήση λατινικού αλφαβήτου προκύπτει από αναγκαιότητα και την έμφυτη δυσκολία αναπαράστασης του πρωτότυπου αλφαβήτου, ενώ στο δεύτερο προκύπτει από θέληση για ευκολία. Συγκεκριμένα:

- Τα Κινέζικα αποτελούν γλώσσα η οποία εμπεριέχει πάνω από 20,000 χαρακτήρες. Κάτι τέτοιο ίσως να μην αποτελεί πρόβλημα κατά την φυσική τους γραφή, όμως παρουσιάζει ανυπέρβατα προβλήματα στην επικοινωνία μέσω της τεχνολογίας. Πράγματι, η επιλογή ενός επιθυμητού χαρακτήρα σε ένα τέτοιο πλήθος χαρακτήρων, με οποιαδήποτε ομαδοποίηση, είναι εξαιρετικά χρονοβόρα και καθιστά την επικοινωνία μέσω τεχνολογικών μέσων πρακτικά αδύνατη. Έτσι, εφευρέθηκε ένα σύστημα όπου τα Κινέζικα γράφονται με λατινικούς χαρακτήρες, με βάση το φωνητικό αλφάβητο, και έπειτα εκτελείται μια διαδικασία transliteration έτσι ώστε να μετατραπεί η επιθυμητή λέξη, από τη φωνητική γραφή της στο λατινικό αλφάβητο, στην παραδοσιακή, Κινέζικη γραφή.
- Τα ελληνικά αποτελούν, επίσης χαρακτηριστικό παράδειγμα όπου πηγάζει οργανικά ένα πρόβλημα transliteration. Πράγματι, για λόγους ταχύτητας και ευκολίας, είναι πολύ συχνό σε μηνύματα, online messaging, και emails, επιλέγεται η γραφή ελληνικών με το λατινικό αλφάβητο, έναντι της χρήσης του ελληνικού αλφαβήτου (η συγκεκριμένη γραφή είναι γνωστή ως greeklish). Σε τέτοιες περιπτώσεις θα ήταν επιθυμητό να γίνεται αυτόματα η μετατροπή στο ελληνικό κείμενο, διατηρώντας παράλληλα, την ευκολία συγγραφής.

Στο Σχήμα 5.1 παρουσιάζουμε ένα συμβολικό παράδειγμα ενός transliteration. Συγκεκριμένα, παρουσιάζεται ένα transliteration μιας πρότασης γραμμένης με το λατινικό αλφάβητο, στην αντίστοιχη πρόταση γραμμένη με το ελληνικό αλφάβητο.



Σχήμα 5.1: Transliteration από ένα αλφάβητο α (λατινικό) σε ένα αλφάβητο β (ελληνικό).

Τα δεδομένα που χρησιμοποιήσαμε στα πειράματα προέρχονται από μια συλλογή email όπου ήταν διαθέσιμο το αρχικό κείμενο, γραμμένο σε greeklish, καθώς και η αντίστοιχη ελληνική ‘μετάφραση’. Αρχικά δημιουργήθηκε ένα δίκτυο το οποίο εμπεριέχει τις απαραίτητες καταστάσεις για όλες τις ελληνικές λέξεις που απαντώνται σε ένα δοσμένο λεξικό, καθώς και ενδιάμεσες καταστάσεις απαραίτητες για τις μεταβάσεις στις τελικές λέξεις. Έπειτα, το δίκτυο αυτό εμπλουτίστηκε με όλες τις απαραίτητες καταστάσεις για τις αγγλικές λέξεις που απαντώνται σε ένα δοσμένο αγγλικό λεξικό. Ο λόγος την επιλογή της μοντελοποίησης αμιγώς αγγλικών λέξεων είναι ότι, συχνά, στα greeklish, χρησιμοποιούνται αυτούσιες αγγλικές λέξεις, τις οποίες θα θέλαμε να συμπεριλάβουμε στο δίκτυο έτσι ώστε να παραμείνουν αναλλοίωτες μετά το πέρας της μετατροπής, εάν αυτές απαντώνται στο κείμενο. Έπειτα, χρησιμοποιώντας

τα δεδομένα, υπολογίσαμε τις πιθανότητες μετατροπής των λατινικών συμβόλων σε ελληνικά. Κάτι τέτοιο παρουσίασε κάποια προβλήματα, καθώς υπάρχει πλήθος συμβόλων, τόσο λατινικά όσο και ελληνικά, τα οποία αντιστοιχούν σε πλήθος συμβόλων στο άλλο αλφάβητο. Χαρακτηριστικό παράδειγμα αποτελούν:

- το γράμμα ελληνικού αλφαβήτου 'Ψ', το οποίο αντιστοιχίζεται στην ακολουθία συμβόλων 'PS' στο λατινικό αλφάβητο.
- το γράμμα λατινικού αλφαβήτου 'D', το οποίο αντιστοιχίζεται στην ακολουθία συμβόλων 'NT' στο ελληνικό αλφάβητο.

Έγινε προσπάθεια να αναγνωριστούν αυτές οι περιπτώσεις, και να μοντελοποιηθούν καταλλήλως. Μετά την μοντελοποίηση υπήρχαν και μερικές αντιστοιχίσεις οι οποίες δεν επαφίενται τις πραγματικότητας. Αυτές αποκόπηκαν, αυτόματα, με βάση έναν ποσοστιαίο κανόνα εμφάνισης των αντιστοιχίσεων στο σύνολο των δεδομένων.

Πέραν λεπτομερειών αναφορικά με τον καθορισμό των μεγεθών του συστήματος, έγιναν κάποιες σχεδιαστικές επιλογές έτσι ώστε να μπορέσει να λειτουργήσει το σύστημα, τις οποίες κρίνουμε σκόπιμο να αναφέρουμε:

- Υποθέσαμε ότι κάθε ένα από τα σύμβολα εισόδου έχει τη δυνατότητα να παρατηρηθεί σε κάθε κατάσταση, ακόμα και αν τα δεδομένα δεν υποστήριζαν κάτι τέτοιο. Ο λόγος γι' αυτή την επιλογή βασίζεται στο ότι θέλαμε να εξετάζονται περισσότερες καταστάσεις κατά την εκτέλεση του αλγορίθμου Viterbi με αποκοπή, και συνεπώς να γίνεται ένας πιο ουσιώδης αποκλεισμός πιθανών μονοπατιών. Επιπλέον, αυτή η επιλογή μπορεί να θεωρηθεί ότι μοντελοποιεί θόρυβο στην παρατήρηση των συμβόλων.
- Υποθέσαμε ότι κάθε μετάβαση μεταξύ κάποιου ζεύγους καταστάσεων είναι πιθανή, ακόμα και αν τα δεδομένα δεν υποστήριζαν κάτι τέτοιο. Αυτή η θεώρηση έγινε καθώς, διαφορετικά, τα μονοπάτια τα οποία εξετάζονταν σε κάθε χρονικό πλαίσιο του αλγορίθμου ήταν λίγα σε αριθμό, το οποίο δεν επέτρεπε την υπογράμμιση της λειτουργίας του αλγορίθμου. Επιπλέον, αυτή η επιλογή μπορεί να θεωρηθεί ότι 'συγχωρεί' λάθη τα οποία πιθανώς να έγιναν σε προηγούμενα βήματα, καθώς επιτρέπει να αναδυθεί η, εν τέλη, βελτιστότητα των μονοπατιών.
- Τέλος, υποθέσαμε ότι κάθε κατάσταση έχει τη δυνατότητα να είναι αρχική κατάσταση. Αυτή η επιλογή έγινε για υπολογιστικούς λόγους, έτσι ώστε να μην απειρίζονται τα αρχικά κόστη.

Οι επιλογές αυτές έγιναν επειδή το αρχικό σύστημα ήταν μοντελοποιημένο αμιγώς μέσω *automata*. Αυτό είχε ως αποτέλεσμα ότι δεν γινόταν προσπάθεια εύρεσης βέλτιστου μονοπατιού σε κάθε βήμα, αλλά εξεταζόταν, απλώς, η ύπαρξη αποδεκτού μονοπατιού συνολικά, και γινόταν χρήση των πιθανοτήτων μονάχα για την επιλογή του 'πιθανότερου' σε περίπτωση που υπήρχαν πολλά αποδεκτά μονοπάτια. Η μετατροπή του προαναφερθέντος συστήματος σε ένα μοντέλο που θα μπορούσαμε να χρησιμοποιήσουμε τον αλγόριθμο Viterbi με αποκοπή ουσιωδώς είχε αρκετούς περιορισμούς, οι οποίοι οδήγησαν στην αναγκαιότητα των παραπάνω επιλογών, αλλά και σε περιορισμούς ως προς την κλίμακα του μετέπειτα πειραματισμού.

Παρ' όλα αυτά, το πείραμα που εκτελέσαμε αφορούσε τον εντοπισμό της πιθανότερης 'μετάφρασης' μιας δοσμένης λέξης μέσω του συστήματος transliteration από λατινικούς σε ελληνικούς χαρακτήρες. Στα πλαίσια του πειράματος εξετάσαμε διαφορετικές τιμές για την παράμετρο χαλάρωσης θ , και καταγράψαμε τις μέσες τιμές των μετρικών της κανονικοποιημένης εντροπίας ϵ και του κανονικοποιημένου όγκου ν . Αποτελέσματα από τον πειραματισμό για δύο συγκεκριμένες λέξεις παρουσιάζονται στον Πίνακα 5.1.

Transliteration από λατινικό αλφάβητο σε ελληνικό αλφάβητο						
input	θ	time (s)	ε	ν	min	max
\ELLIPEIS\	0	89.5	0.0248	0	1	1
	5	121.7	0.0018	1.558	1	1444
	10	201.9	0.0013	2.094	101	3829
	15	533.0	0.0001	1.630	5145	10333
	∞	580.3	0.0001	0	10333	10333
\ALLA\	0	77.6	0.0616	0	1	1
	5	93.3	0.0039	1.435	1	1215
	10	175.2	0.0026	2.072	153	5431
	15	481.8	0.0003	1.765	7088	14246
	∞	562.9	0.0002	0	14246	14246

Πίνακας 5.1: Χρόνος εκτέλεσης, τιμές μετρικών ν και ε , ελάχιστος και μέγιστος αριθμός επιζησάντων καταστάσεων κατά μήκος χρονικών πλαισίων για διαφορετικές τιμές της παραμέτρου χαλάρωσης θ για δύο διαφορετικές λέξεις εισόδου.

5.1.1 Συμπεράσματα

Ακόμα και σε αυτό το σύντομο πείραμα είμαστε σε θέση να κάνουμε μερικές καίριες παρατηρήσεις βασιζόμενοι στον Πίνακα 5.1, αναλογιζόμενοι τα αποτελέσματα της Ενότητας 4.4. Συγκεκριμένα:

- Παρατηρούμε ότι για διαδοχικές τιμές της παραμέτρου χαλάρωσης θ όπου ελαχιστοποιείται η μεταβολή της μετρικής της κανονικοποιημένης εντροπίας ε έχουμε μικρές διαφορές στις καταγεγραμμένες τιμές για τον υπολογιστικό χρόνο και για το πλήθος εξεταζόμενων καταστάσεων. Αυτό έρχεται σε πλήρη συμφωνία με την ανάλυση που κάναμε στην Ενότητα 4.4, καθώς υποδηλώνει ότι θα υπήρχε μικρή ‘έκπληξη’ από την μεταβολή μεταξύ των δύο τιμών της παραμέτρου, και άρα σε εκείνο το εύρος τιμών εντοπίζεται ένα τοπικό βέλτιστο της παραμέτρου χαλάρωσης θ .
- Παρατηρούμε ότι για διαδοχικές τιμές της παραμέτρου χαλάρωσης θ όπου μεγιστοποιείται η τιμή της μετρικής του κανονικοποιημένου όγκου ν (και όπου η αντίστοιχη μεταβολή στη μετρική ε είναι μικρή) έχουμε καλές αποδόσεις του αλγορίθμου αναφορικά με το ισοζύγιο ταχύτητας/επίδοσης. Συγκεκριμένα, εξετάζοντας τις τιμές όπου η παράμετρος θ ισούται με 5 και 10 αντίστοιχα, παρατηρούμε ότι δεν παρουσιάζεται σημαντική μεταβολή στην μετρική ε . Όμως, για την τιμή 10 παρατηρούμε ότι μεγιστοποιείται η μετρική ν . Παράλληλα, εξετάζοντας τον αριθμό των εξεταζόμενων καταστάσεων για τις δύο επιλογές της παραμέτρου θ , βλέπουμε ότι για την τιμή $\theta = 5$ ο ελάχιστος αριθμός καταστάσεων που εξετάζεται σε κάποιο πλαίσιο είναι 1. Αυτό συνεπάγεται ότι, στο συγκεκριμένο πλαίσιο, εξετάζεται μονάχα το βέλτιστο μονοπάτι, και συνεπώς υπάρχει σημαντική πιθανότητα να αποκοπεί η μετέπειτα βέλτιστη διαδρομή. Έτσι, σε αυτές τις περιπτώσεις η μεγιστοποίηση της μετρικής του κανονικοποιημένου όγκου ν μας οδηγεί στη βέλτιστη επιλογή της παραμέτρου χαλάρωσης.

Αξίζει να αναφερθεί ότι στα παραπάνω δεν γίνεται αναφορά στην ακρίβεια της ορθής μετατροπής των λέξεων. Ο λόγος γι’ αυτό είναι ότι, με τη συγκεκριμένη μοντελοποίηση, η αναμενόμενη μετατροπή ήταν η βέλτιστη για κάθε τιμή της παραμέτρου χαλάρωσης θ , και απλώς μεταβαλλόταν ο χρόνος μέχρι να εντοπιστεί, καθώς και μεταβαλλόταν το πλήθος των καταστάσεων που εξεταζόταν. Έγινε εκτεταμένη προσπάθεια να επιλεγούν κατάλληλα

οι τιμές των πιθανοτήτων παρατήρησης και μεταβάσεων έτσι ώστε να παρατηρούνται, περιορισμένα σε αριθμό, 'λάθη' μετατροπής, όμως αυτό αποδείχθηκε εξαιρετικά δύσκολο, καθώς ο αριθμός των καταστάσεων είναι συνταρακτικά μεγάλος και ο ακριβής προσδιορισμός των τιμών που να επιτρέπουν τα λάθη, αλλά με μέτρο, είναι εξαιρετικά περίπλοκος.

5.2 Πειράματα με προσομοίωση σε IP Spoofing

Εξετάζοντας τα αποτελέσματα του πειραματισμού της Ενότητας 5.1 ενισχύσαμε την διαίσθησή που παρουσιάσαμε αρχικά στην Ενότητα 4.4. Έχοντας επιβεβαιώσει την αρχική μας διαίσθηση για τη λειτουργία των μετρικών ε και ν , θα προχωρήσουμε σε περαιτέρω πειραματισμό, με σκοπό την πολύπλευρη εξέταση των μετρικών, τόσο αναφορικά με την χρησιμότητά τους σε διαφορετικές αναλύσεις, αλλά κυρίως σε διαφορετικά πεδία εφαρμογής. Ένας πολύς συχνός τρόπος για μελέτη και πειραματισμό, τόσο στον τομέα της Επιστήμης των Υπολογιστών, αλλά και στον τομέα των Τηλεπικοινωνιών, είναι οι προσομοιώσεις. Η πιο συνηθισμένη μορφή προσομοίωσης είναι οι προσομοιώσεις Monte Carlo ([9], [17], [30]). Σε αυτές οι ερευνητές βασίζονται στο γεγονός ότι η τυχαία παραγωγή δειγμάτων θα οδηγήσει στην επίλυση ντετερμινιστικών προβλημάτων. Γενικότερα, οι προσομοιώσεις αποτελούν συχνά το αντικείμενο εφαρμογών για τρεις κυρίως λόγους:

- Επιτρέπουν τον γρήγορο πειραματισμό. Το γεγονός αυτό προκύπτει από το ότι οι είσοδοι είναι συνήθως μικρά αρχεία κειμένου, σε αντίθεση με τα πειράματα σε πραγματικά δεδομένα όπου οι είσοδοι μπορεί να είναι εικόνες μεγάλων διαστάσεων. Επιπλέον, τα πειράματα επιδέχονται ταχύτατες και εκτενείς αλλαγές παραμέτρων, το οποίο επιτρέπει στους ερευνητές να εξετάσουν πληθώρα ιδεών, και να εντοπίσουν γρήγορα υποσχόμενες τιμές παραμέτρων, χωρίς την ανάπτυξη σύνθετων συστημάτων.
- Συχνά υπάρχει έλλειψη διαθέσιμων δεδομένων, ή είναι πιθανό η τροποποίηση των δεδομένων έτσι ώστε να μπορεί να εφαρμοστεί η εκάστοτε ιδέα να είναι εξαιρετικά δύστροπη και χρονοβόρα. Συνεπώς, σε τέτοιες περιπτώσεις οι ερευνητές χρειάζεται να αναλωθούν σε εκτενή επεξεργασία δεδομένων αποσκοπώντας στην εξέταση των ιδεών τους, αντί να επικεντρώνονται στην ανάπτυξη και επέκτασή τους.
- Τέλος, συχνά, τα πιθανά πεδία εφαρμογής των ιδεών διαφέρουν σημαντικά από τα πεδία έρευνας των ερευνητών. Αυτό έχει ως αποτέλεσμα ότι η ορθή αξιολόγηση των ιδεών μπορεί να απαιτεί συγκεκριμένες γνώσεις και εμπειρία στον τομέα, τις οποίες ο ερευνητής δεν κατέχει. Επομένως, η αξιολόγηση της μεθόδου μέσω της σύγκρισης με τις τρέχουσες ιδέες του πεδίου θα απαιτούσε μήνες ενασχόλησης, χωρίς να του προσφέρουν στον ερευνητή αξιοποιήσιμα εφόδια για την μετέπειτα ερευνητική του πορεία.

Στα πλαίσια αυτής της διπλωματικής επιλέξαμε να επιστρατεύσουμε την τεχνική των προσομοιώσεων Monte Carlo για την περαιτέρω εξέταση των ιδεών της Ενότητας 4.4, για λόγους οι οποίοι επαφίενται και στους τρεις παραπάνω άξονες. Συγκεκριμένα, οι κύριοι λόγοι στους οποίους βασίσαμε την επιλογή μας είναι οι εξής:

- Τα δίκτυα τα οποία πρέπει να δημιουργηθούν για να εξεταστούν πραγματικά δεδομένα έχουν εξαιρετικά μεγάλο όγκο. Χαρακτηριστικό παράδειγμα είναι τα δίκτυα που χρειάζεται να κατασκευαστούν για την αυτόματη αναγνώριση φωνής. Συγκεκριμένα, αυτά τα δίκτυα πρέπει να έχουν καταστάσεις για όλες τις πιθανές λέξεις μιας γλώσσας, καθώς και μεταβατικές καταστάσεις ώστε να δύναται ο αλγόριθμος να αναγνωρίζει τις λέξεις. Ο όγκος των δικτύων συνεπάγεται ότι δεν είναι δυνατή η εκτέλεση των αλγορίθμων σε προσωπικό υπολογιστή. Για την αποφυγή αυτού του προβλήματος θα μπορούσαμε

να εξετάσουμε κάποιο υποσύνολο του δικτύου, και έπειτα να προσπαθήσουμε να γενικεύσουμε τα αποτελέσματα για το υπόλοιπο δίκτυο. Όμως, από αυτή την παραδοχή πηγάζει το πρόβλημα ότι όταν εξετάζεται ένα αυθαίρετο, και όχι προσεκτικά κατασκευασμένο, υποσύνολο του δικτύου τότε η αποκοπή είναι συχνά τετριμμένη, καθώς το βέλτιστο μονοπάτι έχει σημαντικά μικρότερο κόστος από τα υπόλοιπα (ή μπορεί να είναι και το μοναδικό δυνατό). Μέσω των προσομοιώσεων μπορέσαμε να ορίσουμε ένα διαχειρίσιμο μέγεθος δικτύου, το οποίο μας επέτρεψε να διατηρήσουμε την φυσική ερμηνεία, και παράλληλα να υπογραμμίσουμε τις ιδιότητες του αλγορίθμου που μας ενδιέφεραν περισσότερο.

- Τα δύο πεδία στα οποία απαντάται με μεγαλύτερη συχνότητα ο αλγόριθμος Viterbi είναι η αναγνώριση φωνής και η βιοπληροφορική, και συγκεκριμένα στα προβλήματα της αυτόματης αναγνώρισης φωνής και του DNA sequencing. Αυτά τα δύο πεδία αποτελούν αντικείμενα με πολύ μεγάλο ενδιαφέρον από εταιρίες αλλά και ερευνητικές ομάδες πανεπιστημίων. Αυτό συνεπάγεται ότι τα ελεύθερα διαθέσιμα δεδομένα είναι αρκετά περιορισμένα, ή ακόμα και μη διαθέσιμα, καθώς κάθε είδους ερευνητική ομάδα είναι διστακτική στο να μοιραστεί τα δεδομένα τα οποία έχει στη διάθεσή της. Έτσι, τα διαθέσιμα σύνολα δεδομένων αποτελούνται από πολύ λίγες παρατηρήσεις, και για συγκεκριμένα υποσύνολα του προβλήματος (για παράδειγμα, για την αυτόματη αναγνώριση φωνής υπάρχει διαθέσιμο σύνολο δεδομένων από εκφωνήσεις των ψηφίων από 0 έως 9 από 7 διαφορετικούς άντρες).
- Αναφορικά με το DNA sequencing, η πρόωμη μας επαφή με τον τομέα της βιοπληροφορικής είναι πολύ περιορισμένη. Κατ'επέκταση η υλοποίηση των ιδεών της Ενότητας 4.4 στο προκείμενο πρόβλημα, πέραν του χρονικού παραβόλου απαιτούμενου για την εξοικείωση μας με τις ιδιαιτερότητες του πεδίου και του προβλήματος, θα ήταν ελλιπής καθώς μέσω της αξιολόγησης δεν θα ήταν εμφανές εάν η απόδοση ευθύνεται στους έμφυτους περιορισμούς του αλγορίθμου και της προτεινόμενης ιδέας, ή στην περιορισμένη επαφή μας με το πεδίο της βιοπληροφορικής..

Έχοντας θίξει τα γενικά οφέλη του πειραματισμού μέσω προσομοίωσης, και έχοντας αναλύσει τους συγκεκριμένους λόγους που οδήγησαν στην επιλογή των προσομοιώσεων Monte Carlo για την ολοκλήρωση της προκείμενης διπλωματικής εργασίας, χρειάζεται να ακολουθήσει ο σχεδιασμός των πειραμάτων. Συγκεκριμένα, πρέπει να επιλέξουμε ποια ακριβώς θα είναι η δομή τους, αλλά παράλληλα να καθορίσουμε και τον χαρακτήρα τους, δηλαδή τι, ακριβώς, θα θέλαμε το κάθε πείραμα να υπογραμμίσει. Θα θέλαμε μέσω των πειραμάτων να υπογραμμίσουμε τις εξής δυνατότητες των μετρικών ϵ και ν :

- Θα θέλαμε να σχεδιάσουμε ένα πείραμα το οποίο να υπογραμμίζει τη δυνατότητα των μετρικών, και κυριότερα της μετρικής ϵ , να προβλέψουν το 'κέρδος' στην παρατηρούμενη ακρίβεια για διαφορετικές τιμές της παραμέτρου χαλάρωσης. Συγκεκριμένα, θα θέλαμε οι τιμές της ακρίβειας για τις διαφορετικές τιμές της παραμέτρου χαλάρωσης να έχουν σημαντική απόκλιση όταν παρατηρείται και αντίστοιχα μεγάλη μεταβολή στις τιμές της μετρικής ϵ , όπως υπονοείται από την Ενότητα 4.4.
- Επίσης, θα θέλαμε στα πειράματα να είναι εμφανές το γεγονός ότι οι μετρικές οδηγούν στην βέλτιστη επιλογή της παραμέτρου χαλάρωσης. Δηλαδή, για ίδιες καταγεγραμμένες τιμές ακρίβειας, θα θέλαμε οι μετρικές να οδηγούν στην επιλογή της τιμής παραμέτρου χαλάρωσης η οποία έχει το μικρότερο κόστος (το οποίο, συνήθως, σημαίνει την μικρότερη χρονική πολυπλοκότητα).

- Θα θέλαμε να εξετάσουμε τη δυνατότητα των μετρικών στον αυτόματο προσδιορισμό της παραμέτρου χαλάρωσης θ . Ο υπολογισμός των βέλτιστων υπερπαραμέτρων για ένα πρόβλημα αποτελεί ένα από τα μελανά σημεία της μηχανικής μάθησης, και μια από τις πιο ανιαρές διαδικασίες με τις οποίες χρειάζεται να καταπιαστούν οι ερευνητές. Συνεπώς, ο αυτόματος υπολογισμός της παραμέτρου, χρησιμοποιώντας τις μετρικές ε και ν έχει μεγάλη βαρύτητα.

Συμπεριλαμβάνοντας στη διαδικασία σχεδιασμού τους παραπάνω προβληματισμούς/στόχους, καταλήξαμε στο σχεδιασμό τριών διακριτών πειραμάτων. Κρίνουμε ότι τα πειράματα αυτά επιτυγχάνουν, σε επαρκή βαθμό, τα παραπάνω ζητούμενα, και συνεπώς υπογραμμίζουν τις ζητούμενες ιδιότητες των μετρικών ε και ν . Συγκεκριμένα, τα πειράματα που σχεδιάσαμε είναι τα εξής:

- Το Πείραμα 1 αποτελείται από την εκτέλεση τον αλγόριθμο Viterbi με αποκοπή για διαφορετικές τιμές της παραμέτρου χαλάρωσης θ σε ένα δίκτυο με ένα δοσμένο αριθμό καταστάσεων, και την καταγραφή των τιμών των μετρικών ε και ν . Οι καταστάσεις παράγουν δείγματα με βάση μια επιλεγμένη κατανομή, και μια από τις καταστάσεις ακολουθεί διαφορετική κατανομή από τις υπόλοιπες. Το πείραμα αποσκοπεί να υπογραμμίσει τη δυνατότητα των μετρικών να εντοπίσουν μεταβολές στην ακρίβεια, καθώς και το γεγονός ότι οδηγούν στην επιλογή της παραμέτρου χαλάρωσης με το μικρότερο κόστος.
- Το Πείραμα 2 αποτελείται από την εκτέλεση του αλγορίθμου Viterbi με αποκοπή για διαφορετικές τιμές της παραμέτρου χαλάρωσης θ στο ίδιο δίκτυο με αυτό του Πειράματος 1. Τα πειράματα διαφέρουν στο ότι στο Πείραμα 2 επιτρέπουμε η κατάσταση που ακολουθεί διαφορετική κατανομή να αλλάζει μεταξύ των διαδοχικών χρονικών πλαισίων, αντίθετα με το Πείραμα 1 όπου θεωρήσαμε ότι η κατάσταση διαφορετικής κατανομής παραμένει σταθερή μεταξύ διαδοχικών χρονικών πλαισίων.
- Το Πείραμα 3 αποτελείται από μια συλλογή επιμέρους πειραμάτων στα οποία, με βάση τις καταγεγραμμένες τιμές των μετρικών ε και ν σε κάθε χρονικό πλαίσιο, αποσκοπούμε στην αυτόματη μεταβολή της παραμέτρου χαλάρωσης έτσι ώστε να ελαχιστοποιήσουμε την σπατάλη των υπολογιστικών πόρων, επιτυγχάνοντας παράλληλα ικανοποιητικά ποσοστά ακρίβειας. Το πείραμα αποσκοπεί να αντιμετωπίσει τον τρίτο από τους προβληματισμούς που εισήγαμε παραπάνω.

5.2.1 Θεωρητικό Υπόβαθρο

Προτού προβούμε στα χωρία των πειραμάτων, κρίνουμε σκόπιμο να αφιερώσουμε μια σύντομη ενότητα στην επεξήγηση του πλαισίου γύρω από το οποίο έγινε η ανάπτυξη των πειραμάτων. Στον τομέα των τηλεπικοινωνιών, ειδικά στη σημερινή εποχή με την έξαρση του διαδικτύου και των ψηφιακών δικτύων, έχει προκύψει το πρόβλημα της ασφάλειας των δικτύων. Συγκεκριμένα, είναι καθημερινό φαινόμενο κάθε λογής κακόβουλοι χρήστες να προσπαθούν, κάνοντας μεγάλο αριθμό αιτημάτων, να διαταράξουν την ομαλή λειτουργία ενός δικτύου. Σε τέτοιες περιπτώσεις το δίκτυο, προσπαθώντας να αντεπεξέλθει στον μεγάλο αριθμό αιτημάτων του κακόβουλου χρήστη, αδυνατεί να ανταποκριθεί στα αιτήματα των έγκυρων χρηστών. Συνεπώς, η αυτόματη ανίχνευση τέτοιων κακόβουλων χρηστών αποτελεί πρόβλημα με μεγάλο ενδιαφέρον, όπως επίσης και ο εμποδισμός των αιτημάτων αυτών των χρηστών. Ερευνητές έχουν προτείνει διάφορους τρόπους για τον αυτόματο εντοπισμό αυτών των χρηστών. Συγκεκριμένα, στα [54] και [55] οι ερευνητές προσπαθούν σε διαδοχικά βήματα να συγκεκριμενοποιήσουν την πηγή των αιτημάτων κάνοντας χρήση της πληροφορίας για το σύνολο των αιτημάτων που πηγάζουν από δέσμες διευθύνσεων IP. Έχοντας στη διάθεσή τους αυτή την πληροφορία, εκλεπτίζουν,

σε κάθε βήμα, τη δέσμη διευθύνσεων που θεωρούν ως πηγή, οδηγώντας διαδοχικά στον κακόβουλο χρήστη. Παρατηρούμε ότι η παραπάνω διεργασία εμφανίζει σημαντικές ομοιότητες με την γενικότερη διαδικασία της αποκοπής, και συνεπώς το πρόβλημα δύναται να μοντελοποιηθεί επαρκώς στη μεθοδολογία του αλγορίθμου Viterbi με αποκοπή.

Συχνά στον τομέα των τηλεπικοινωνιών τα αιτήματα που εκτελούν οι χρήστες σε μια υπηρεσία ή σε ένα δίκτυο μοντελοποιούνται με κατανομές Poisson ([2], [46]). Η κατανομή Poisson καθορίζεται από την παράμετρο λ , που εκφράζει τον μέσο αριθμό αιτημάτων μέσα σε ένα χρονικό παράθυρο. Εναλλακτικά, στα συστήματα και τις ουρές αναμονής η κατανομή Poisson χρησιμοποιείται για να μοντελοποιήσει τον χρόνο που αναμένεται να περιμένει ένας χρήστης έως ότου εξυπηρετηθεί. Σε αυτό το πλαίσιο, η παράμετρος λ της κατανομής εκφράζει τον μέσο χρόνο αναμονής. Η πιθανότητα να παρατηρηθούν ακριβώς κ αιτήματα μέσα στο χρονικό παράθυρο (ή, αντίστοιχα, ένας χρήστης να περιμένει ακριβώς κ λεπτά μέχρι να εξυπηρετηθεί) δίνεται από την Εξίσωση (5.1):

$$P[X = \kappa] = e^{-\lambda} \frac{\lambda^{\kappa}}{\kappa!} \quad (5.1)$$

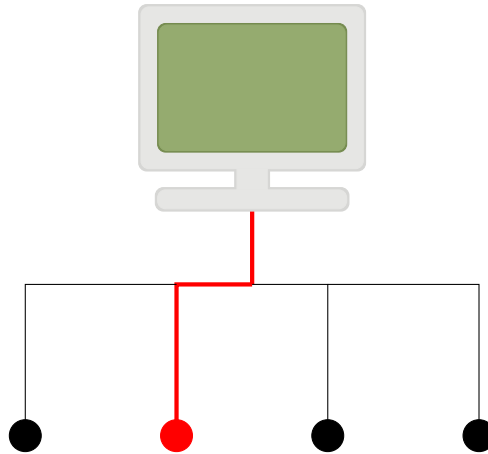
όπου X είναι η τυχαία μεταβλητή που μοντελοποιεί τον αριθμό των αιτημάτων στο χρονικό παράθυρο (ή τον χρόνο αναμονής στην περίπτωση των ουρών αναμονής). Αξίζει να αναφέρουμε μερικά σημαντικά χαρακτηριστικά της κατανομής. Συγκεκριμένα, ισχύουν τα παρακάτω:

- Η κατανομή είναι διακριτή: τα δείγματα που προκύπτουν από δειγματοληψία της κατανομής έχουν διακριτές τιμές.
- Η μέση τιμή μ της κατανομής είναι ίση με την παράμετρο της κατανομής, λ .
- Η διασπορά σ^2 της κατανομής είναι ίση με την παράμετρο της κατανομής, λ . Αυτή η ιδιότητα παρουσιάζει μεγάλο ενδιαφέρον, αφ' ενός επειδή η διασπορά είναι ίση με τη μέση τιμή, και αφ' εταίρου επειδή η διασπορά έχει μεγάλη τιμή.
- Η κατανομή έχει την ιδιότητα της *έλλειψης μνήμης*. Αυτό, πρακτικά, σημαίνει ότι εάν ένας χρήστης έχει αναμένει ήδη χρόνο τ , τότε η πιθανότητα να περιμένει συνολικό χρόνο $t + \tau$ είναι ίση με την πιθανότητα να ανέμενε χρόνο t . Μαθηματικά το παραπάνω εκφράζεται:

$$P[X > t + \tau | X \geq \tau] = P[X > t] \quad (5.2)$$

Διαφορετικά, ο χρόνος που έχει αναμείνει ήδη ο χρήστης δεν επηρεάζει τον εκτιμώμενο αναμενόμενο χρόνο. Εάν ο μέσος χρόνος αναμονής είναι $\lambda = 5$ λεπτά, τότε η πιθανότητα ένας χρήστης που έχει αναμείνει ήδη 40 λεπτά να αναμείνει, συνολικά, 45 λεπτά, και ένας νέος χρήστης στο δίκτυο να αναμείνει συνολικά 5 λεπτά, είναι ίσες.

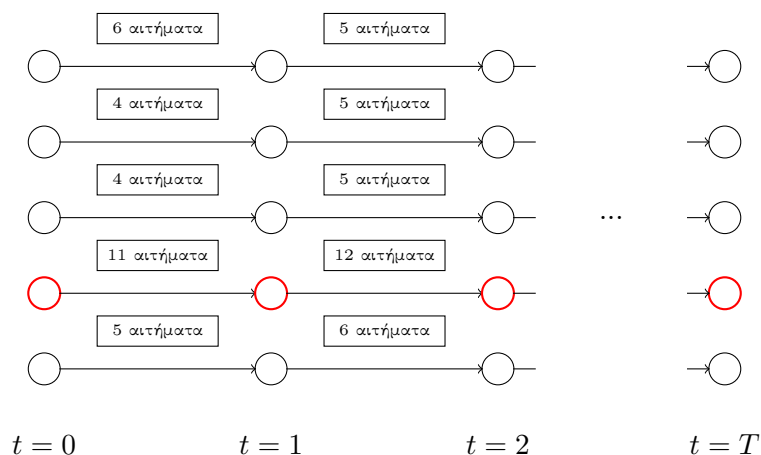
Οι παραπάνω ιδιότητες αποτελούν λόγους για τους οποίους επιλέγεται η κατανομή Poisson σε αυτού του είδους τα προβλήματα. Συγκεκριμένα, αφ' ενός θέλουμε να παράγουμε διακριτό αριθμό αιτημάτων έτσι ώστε να υπάρχει συνέπεια με τη μοντελοποίηση. Αφ' εταίρου, η μεγάλη διασπορά της κατανομής επιφέρει και μεγάλη διασπορά στον αριθμό των αιτημάτων, κάτι το οποίο είναι θεμιτό καθώς μοντελοποιεί επαρκώς την τυχαιότητα, αλλά παράλληλα επιτρέπει για πιο ενδιαφέρουσα μελέτη (καθώς, εάν η διασπορά είναι μικρή, ο εντοπισμός του κακόβουλου χρήστη είναι σημαντικά ευκολότερη). Τέλος, η ιδιότητα της έλλειψης μνήμης είναι μια από τις σημαντικότερες ιδιότητες της κατανομής Poisson, και έχει δειχθεί πειραματικά ότι μοντελοποιεί ικανοποιητικά τα αιτήματα του διαδικτύου. Συνεπώς, τα πειράματα που θα εκτελέσουμε θα χρησιμοποιούν κυρίως προσομοιώσεις με την κατανομή Poisson. Εξαίρεση σε αυτόν τον κανόνα θα αποτελέσουν κάποια υποπειράματα του Πειράματος 3, τα οποία έχουν σκοπό να καθησυχάσουν τον αναγνώστη, παρουσιάζοντας αντίστοιχα αποτελέσματα και σε άλλες κατανομές, και συγκεκριμένα στην *κανονική κατανομή* ([2], [46]).



Σχήμα 5.2: Αιτήματα χρηστών σε έναν κεντρικό υπολογιστή. Με κόκκινο έχει επισημειωθεί ο κακόβουλος χρήστης που εκτελεί αιτήματα με σημαντικά μεγαλύτερο ρυθμό.

5.2.2 Πείραμα 1: Εντοπισμός αποκλίνουσας κατάστασης

Στο πρώτο πείραμα θεωρούμε ότι υπάρχει ένας κακόβουλος χρήστης ο οποίος επιθυμεί να διαταράξει τη λειτουργία του δικτύου, και ο σκοπός μας είναι να τον εντοπίσουμε χωρίς να σπαταλήσουμε άσκοπα τους πόρους του δικτύου. Θα μοντελοποιήσουμε τους χρήστες ως κόμβους που παράγουν δείγματα σύμφωνα με κάποια κατανομή Poisson. Οι αριθμοί που παράγονται ως δείγματα σε κάθε βήμα μοντελοποιούν τον αριθμό των αιτημάτων που έκανε ο χρήστης μέσα στο (αυθαίρετου μεγέθους) χρονικό παράθυρο. Συνεπώς, το πείραμα αποσκοπεί στον εντοπισμό του κακόβουλου χρήστη, χρησιμοποιώντας ως παρατηρήσεις τον αριθμό των αιτημάτων που εκτελεί ο κάθε χρήστης στο εκάστοτε χρονικό πλαίσιο. Ο εντοπισμός θα τελεσθεί κάνοντας χρήση του αλγορίθμου Viterbi με αποκοπή, έτσι ώστε να μην εξετάζονται σε κάθε βήμα όλες οι μεταβάσεις, και συνεπώς να γίνεται οικονομία στους πόρους του δικτύου. Η πειραματική διάταξη παρουσιάζεται σχηματικά στο Σχήμα 5.3.



Σχήμα 5.3: Εντοπισμός αποκλίνουσας κατάστασης: Ο κακόβουλος χρήστης παραμένει σταθερός μεταξύ των χρονικών παραθύρων.

Πειραματισμός

Τα επιμέρους πειράματα που εκτελέστηκαν στο ευρύτερο πλαίσιο του Πειράματος 1 μοντελοποιήθηκαν ως εξής: για διαφορετικές τιμές της παραμέτρου χαλάρωσης θ υπολογίστηκαν οι τιμές των μετρικών που ορίστηκαν στην Ενότητα 4.4 και σημειώθηκαν κάποια μεγέθη με βάση τα οποία αξιολογήθηκε η επίδοση του αλγορίθμου για κάθε διαφορετική τιμή της παραμέτρου θ . Συγκεκριμένα, τα μεγέθη τα οποία σημειώθηκαν κατά την ολοκλήρωση κάθε επιμέρους πειράματος ήταν τα εξής:

- Αρχικά, η εκάστοτε τιμή της παραμέτρου χαλάρωσης θ , έτσι ώστε να είναι δυνατός ο μετέπειτα διαχωρισμός των υπολοίπων μεγεθών.
- Η μέση τιμή της μετρικής ε , που ορίστηκε στην Ενότητα 4.4, κατά μήκος των πλαισίων. Επιλέξαμε τη μέση τιμή, καθώς κρίναμε ότι εκφράζει με μεγαλύτερη ακρίβεια τη συνολική επιρροή της συγκεκριμένης τιμής για την παράμετρο χαλάρωσης.
- Η διαφορά $\Delta\varepsilon$ μεταξύ της μετρικής ε για την τρέχουσα τιμή της παραμέτρου χαλάρωσης θ και της μετρικής ε για την προηγούμενη τιμή της παραμέτρου χαλάρωσης.
- Η μέση τιμή της μετρικής ν , που ορίστηκε στην Ενότητα 4.4 κατά μήκος των πλαισίων. Και πάλι επιλέχθηκε η μέση τιμή καθώς κρίθηκε ότι εκφράζει τη συνολική επιρροή της συγκεκριμένης τιμής της παραμέτρου χαλάρωσης.
- Την *ακρίβεια* του αλγορίθμου Viterbi με αποκοπή στο να εντοπίζει την πραγματική κατάσταση. Υπενθυμίζουμε στον αναγνώστη ότι η ακρίβεια δίνεται από την σχέση:

$$\text{ακρίβεια} = \frac{\#\text{ορθών προβλέψεων}}{\#\text{συνολικών προβλέψεων}} \quad (5.3)$$

- Ένα μέγεθος το οποίο αποκαλούμε *κάλυψη*, και εκφράζει κατά πόσο η *πραγματική κατάσταση* ήταν μέσα στο σύνολο των καταστάσεων που εξέτασε ο αλγόριθμος (ακόμα και αν δεν ήταν αυτή που τελικά επιλέχθηκε από τον αλγόριθμο). Από αυτόν τον ορισμό συνεπάγεται ότι η κάλυψη είναι πάντα μεγαλύτερη (ή ίση) της ακρίβειας.

Επιλέξαμε να παρουσιάζουμε και την διαφορά μεταξύ των μετρικών ε , παρ' όλου που η τιμή της μπορεί να υπολογισθεί από τα διαθέσιμα μεγέθη των πινάκων, προς τη διευκόλυνση του αναγνώστη. Σύμφωνα με το πνεύμα της Ενότητας 4.4, θα αναφερθούμε στη διαφορά των μετρικών ε στην αξιολόγηση της επίδοσης του αλγορίθμου για διαφορετικές τιμές της παραμέτρου χαλάρωσης στην ενότητα των συμπερασμάτων. Σημειώνεται ότι όλες οι τιμές στρογγυλοποιήθηκαν στο πέμπτο δεκαδικό ψηφίο.

Ενδιαφέρον παρουσιάζει, ίσως, στον αναγνώστη η έννοια της κάλυψης. Συγκεκριμένα, πιθανώς να μην είναι εμφανής ο λόγος για τον οποίο χρησιμοποιούμε το εν λόγω μέγεθος. Όμως, η κάλυψη μας επιτρέπει να ποσοτικοποιήσουμε το 'κέρδος' που μας επιφέρει η αύξηση της παραμέτρου χαλάρωσης με έναν διαφορετικό τρόπο. Είναι πιθανό η πραγματική κατάσταση να μην είναι δυνατό να προκύψει ως η βέλτιστη κατάσταση του πλαισίου. Αυτό μπορεί να οφείλεται στην παραμετροποίηση του συστήματος (για παράδειγμα, τις τιμές που επιλέξαμε για τον πίνακα μεταβάσεων), στην διακύμανση των κατανομών, ή του υπάρχοντος θορύβου στο κανάλι (που μπορούμε να θεωρήσουμε ότι απορροφάται ως όρος από την διακύμανση των κατανομών, ή, εναλλακτικά, μπορούμε να τον μοντελοποιήσουμε ρητά ως λευκό γκαουσιανό θόρυβο). Σε μια τέτοια περίπτωση, όσο και αν αυξήσουμε τη παράμετρο χαλάρωσης, η 'πραγματική' κατάσταση δεν θα αναδυθεί ποτέ ως βέλτιστη. Η έννοια της κάλυψης μας προσφέρει έναν εναλλακτικό τρόπο να εκτιμήσουμε την συμπεριφορά του συστήματος: εάν υπάρχει σημαντική απόκλιση μεταξύ της ακρίβειας και της κάλυψης, τότε σπαταλιόνται οι πόροι του δικτύου,

καθώς εξετάζεται μεγαλύτερος αριθμός κόμβων, χωρίς, όμως, αυτό να συνεπάγεται αύξηση των ορθών προβλέψεων. Καταλαβαίνουμε ότι κάτι τέτοιο δεν είναι επιθυμητό, και συνεπώς θα θέλαμε να εκτελεστεί ο σχεδιασμός του συστήματος κατάλληλα.

Για λόγους πληρότητας, πριν προβούμε στην παρουσίαση των αποτελεσμάτων, θα καταγράψουμε τις τιμές των παραμέτρων των πειραμάτων. Συγκεκριμένα:

- Θεωρούμε ότι έχουμε 10 διαφορετικούς χρήστες που κάνουν αιτήματα στο δίκτυο. Τα αιτήματα μοντελοποιούνται με δείγματα από κατανομές Poisson. Οι παράμετροι των κατανομών Poisson τέθηκαν, αυθαίρετα, ίσες με $\lambda = 5$ για τους έγκυρους χρήστες, και λ_d για τον κακόβουλο χρήστη (όπου η τιμή λ_d μεταβάλλεται για τις διαφορετικές εκφάνσεις του πειράματος). Να σημειώσουμε ότι, αντίθετα με την θεωρητική ανάλυση παραπάνω, ο κακόβουλος χρήστης στο πείραμα έχει μικρότερη παράμετρο λ στην κατανομή Poisson. Ο λόγος γι' αυτό είναι ότι ο φορμαλισμός που χρησιμοποιούμε καθ' όλη τη διάρκεια του κειμένου είναι της τροπικής άλγεβρας, και συνεπώς επιζητούμε την ελαχιστοποίηση, και όχι τη μεγιστοποίηση, του συνολικού κόστους. Παράλληλα, για τον διευκόλυνση του αναγνώστη κρίνουμε σκόπιμο να αποφύγουμε τον αρνητικό λογάριθμο των παρατηρήσεων, καθώς τότε οι τιμές απ' ενός δεν θα ήταν ακέραιες, και απ' εταίρου τα μεγέθη δεν θα ήταν άμεσα ερμηνεύσιμα. Συνεπώς, θα ληστεύαμε τις παρατηρήσεις από κάθε είδους φυσική σημασία, και κρίναμε ότι κάτι τέτοιο δεν θα επέφερε διαφορετικά αποτελέσματα, αλλά θα ζημίωνε την κατανόηση του αναγνώστη. Διαβεβαιώνουμε ότι αυτή η επιλογή δεν επηρεάζει καθόλου τα αποτελέσματα.
- Θεωρούμε ότι δεν υπάρχει πρόωμη γνώση στο δίκτυο, και συνεπώς η *a priori* πιθανότητα κάθε χρήστη να είναι ο κακόβουλος είναι ίση με 0.1 (αφού έχουμε 10 χρήστες στο δίκτυο). Επιπλέον, θεωρούμε ότι η πιθανότητα να αλλάξουμε 'κατάσταση' (δηλαδή, ενώ τη χρονική στιγμή t έχουμε επισημειώσει τον Χρήστη 2 ως τον κακόβουλο χρήστη, τη χρονική στιγμή $t+1$ να επισημειώσουμε κάποιον άλλον χρήστη ως τον κακόβουλο) είναι ίση με 0.08, ενώ η πιθανότητα να παραμείνουμε στην ίδια κατάσταση να είναι ίση με 0.18.
- Ο συνολικός αριθμός των χρονικών πλαισίων για τα οποία εκτελέστηκε το εκάστοτε πείραμα τέθηκε, αυθαίρετα, ίσος με 10,000.

Θα προχωρήσουμε, τώρα, στην παρουσίαση των αποτελεσμάτων των επιμέρους πειραμάτων που εκτελέστηκαν στο πλαίσιο του Πειράματος 1. Συγκεκριμένα, θα παρουσιάσουμε τρεις πίνακες οι οποίοι αντιστοιχούν σε διαφορετικές επιλογές για την τιμή της παραμέτρου λ_d της κατανομής Poisson του κακόβουλου χρήστη. Οι τιμές που επιλέξαμε για την παράμετρο τη λ_d είναι 1, 2, και 3 για τους Πίνακες 5.2, 5.3, και 5.4 αντίστοιχα. Τα πειράματα που εκτελέσαμε για κάθε μια από τις επιλογές της παραμέτρου λ_d εξέταζαν διαφορετικές τιμές της παραμέτρου χαλάρωσης θ . Οι τιμές που εξετάστηκαν για την παράμετρο θ σε κάθε επιμέρους πείραμα ήταν 0, 1, 2, 3, και 4.

Στους τίτλους των παρακάτω πινάκων η ποσότητα λ_d εκφράζει την παράμετρο της Poisson κατανομής της αποκλίνουσας κατάστασης. Υπενθυμίζουμε ότι, για το συγκεκριμένο πείραμα, η τιμή της παραμέτρου των Poisson κατανομών των υπολοίπων καταστάσεων είναι $\lambda = 5$.

Συμπεράσματα

Σε αυτή την ενότητα θα αφιερωθούμε στην συνολική ανάλυση των αποτελεσμάτων του πειράματος αποκλίνουσας κατάστασης. Αρχικά, θα αναλύσουμε τα αποτελέσματα αυτούσια, και έπειτα θα σχολιάσουμε πώς συνδέονται με τις παρατηρήσεις που κάναμε στο Κεφάλαιο 4. Από τους Πίνακες 5.2, 5.3, και 5.4 των επιμέρους πειραμάτων μπορούμε να κάνουμε μερικές γενικές παρατηρήσεις για το Πείραμα 1:

Πείραμα αποκλίνουσας κατάστασης ($\lambda_d = 1$)					
θ	ε	$d\varepsilon$	ν	Ακρίβεια	Κάλυψη
0	13.83447×10^{-6}	-	0	0.8467	0.8506
1	19.32590×10^{-6}	5.49143×10^{-6}	0.26164	0.943	0.955
2	18.00068×10^{-6}	-1.32522×10^{-6}	0.71071	0.9416	0.9894
3	5.74471×10^{-6}	-12.25597×10^{-6}	1.13931	0.936	0.9965
4	4.28778×10^{-6}	-1.45693×10^{-6}	1.40026	0.941	0.9996

Πίνακας 5.2: Τιμές μετρικών ν και ε , ακρίβεια και κάλυψη για διαφορετικές τιμές της παραμέτρου χαλάρωσης θ .

Πείραμα αποκλίνουσας κατάστασης ($\lambda_d = 2$)					
θ	ε	$d\varepsilon$	ν	Ακρίβεια	Κάλυψη
0	26.85974×10^{-6}	-	0	0.603	0.6239
1	9.31590×10^{-6}	-17.54384×10^{-6}	0.26825	0.7066	0.8035
2	9.21728×10^{-6}	-0.09862×10^{-6}	0.63932	0.6983	0.91
3	1.25059×10^{-6}	-7.96669×10^{-6}	0.91232	0.6933	0.9638
4	1.49763×10^{-6}	0.24704×10^{-6}	1.05962	0.7057	0.9878

Πίνακας 5.3: Τιμές μετρικών ν και ε , ακρίβεια και κάλυψη για διαφορετικές τιμές της παραμέτρου χαλάρωσης θ .

Πείραμα αποκλίνουσας κατάστασης ($\lambda_d = 3$)					
θ	ε	$d\varepsilon$	ν	Ακρίβεια	Κάλυψη
0	12.90265×10^{-6}	-	0	0.3683	0.3944
1	24.03270×10^{-6}	11.13005×10^{-6}	0.21051	0.4215	0.5953
2	5.87112×10^{-6}	-18.16158×10^{-6}	0.53849	0.416	0.765
3	1.23444×10^{-6}	-4.63668×10^{-6}	0.75673	0.4198	0.8807
4	10.20167×10^{-6}	8.96723×10^{-6}	0.84781	0.424	0.9438

Πίνακας 5.4: Τιμές μετρικών ν και ε , ακρίβεια και κάλυψη για διαφορετικές τιμές της παραμέτρου χαλάρωσης θ .

- Είναι εμφανές ότι όταν η διαφορά μεταξύ της παραμέτρου λ_d της αποκλίνουσας κατάστασης και της παραμέτρου λ των υπολοίπων καταστάσεων είναι μεγάλη, υπάρχει μεγάλο ποσοστό ορθής αναγνώρισης (όπως, για παράδειγμα, παρατηρείται στον Πίνακα 5.2), ενώ όταν το περιθώριο είναι στενό παρατηρούμε σημαντική μείωση στην ακρίβεια (όπως, για παράδειγμα, παρατηρείται στον Πίνακα 5.4).
- Παρατηρούμε ότι η κάλυψη αυξάνεται με την αύξηση της παραμέτρου χαλάρωσης θ , ακόμα και αν δεν παρατηρείται σημαντική αντίστοιχη αύξηση στην τιμή της ακρίβειας. Κάτι τέτοιο είναι αναμενόμενο, όπως υπονοήσαμε και κατά τον ορισμό της κάλυψης, καθώς αυξάνοντας την παράμετρο χαλάρωσης επιτρέπουμε σε περισσότερες καταστάσεις να επιβιώσουν την αποκοπή, και συνεπώς αυξάνεται η πιθανότητα επιβίωσης της πραγματικής κατάστασης. Σημειώνεται ότι για $\theta \rightarrow \infty$ η κάλυψη τείνει στο 1 (και, πρακτικά, για μεγάλες τιμές της παραμέτρου χαλάρωσης η κάλυψη θα είναι ίση με 1).

Προχωρώντας στην εξέταση των αποτελεσμάτων των Πινάκων 5.2, 5.3, και 5.4 συνοπλο-

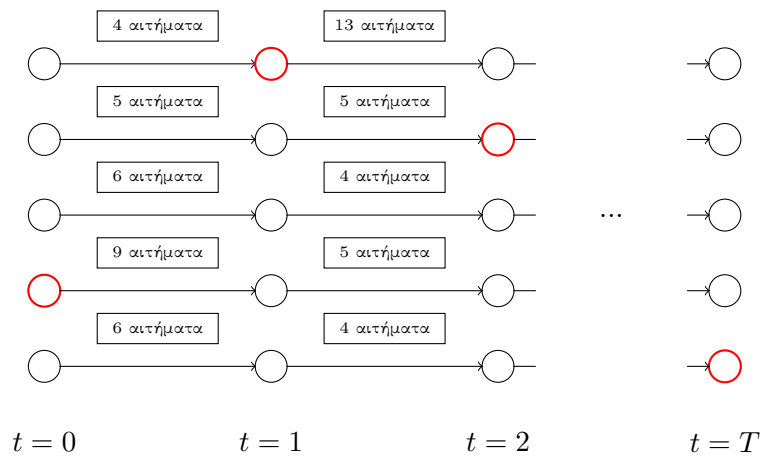
γίζοντας τις παρατηρήσεις της Ενότητας 4.4, έχουμε να σημειώσουμε τα εξής:

- Παρατηρούμε ότι σε κάθε ζεύγος διαδοχικών τιμών της παραμέτρου χαλάρωσης θ που η τιμή του de είναι μικρή, υπάρχει αντίστοιχα μικρή διαφορά στην παρατηρούμενη ακρίβεια μεταξύ των εκτελέσεων του πειράματος, δίνοντας έτσι πειραματική βάση στα διαισθητικά σχόλια της Ενότητας 4.4. Αντίστροφα, παρατηρούμε ότι ζεύγη διαδοχικών τιμών θ για τα οποία παρατηρείται μεγάλη διαφορά στις καταγεγραμμένες τιμές ακρίβειας, παρατηρείται, αντίστοιχα, και μεγάλη διαφορά μεταξύ των τιμών των μετρικών ε .
- Παρατηρούμε ότι για κάποια διαδοχικά ζευγάρια των τιμών της παραμέτρου χαλάρωσης θ παρατηρείται μεγάλη διαφορά στις τιμές των μετρικών ε , χωρίς, όμως, να παρατηρείται αντίστοιχα σημαντική διαφορά στην ακρίβειά τους. Εάν εξετάσουμε, όμως, τις καταγεγραμμένες τιμές για τις αντίστοιχες καλύψεις, παρατηρούμε ότι σε αυτές τις περιπτώσεις υπάρχει σημαντική αύξηση της κάλυψης. Συνεπώς, η σημαντική μεταβολή στις καταγεγραμμένες τιμές της μετρικής ε για διαδοχικές τιμές της παραμέτρου χαλάρωσης θ έχει δύο δυνατές πηγές: είτε προέρχεται από σημαντική αύξηση στην ακρίβεια, είτε προέρχεται από σημαντική αύξηση στην κάλυψη.

Η τελευταία παρατήρηση έχει ενδιαφέρον. Στον εν λόγω πρόβλημα, θα θέλαμε να ελαχιστοποιήσουμε τους πόρους που καταναλώνει το σύστημα, και συνεπώς θα θέλαμε να ελαχιστοποιήσουμε τον αριθμό εξεταζόμενων καταστάσεων, έτσι ώστε να μειώσουμε τον χρόνο απόκρισης, και κατά συνέπεια, να περιορίσουμε την πιθανή ζημιά που μπορεί να προκαλέσει ένας κακόβουλος χρήστης. Όμως, σε άλλα προβλήματα μπορεί να είναι σημαντικότερο να μεγιστοποιήσουμε την πιθανότητα εύρεσης της πραγματικής κατάστασης (πάντα, όμως, με μια ρήτρα στο υπολογιστικό κόστος). Σε αυτές τις περιπτώσεις θα είχε ενδιαφέρον η επιλογή τιμών για την παράμετρο χαλάρωσης οι οποίες, για ίδιες τιμές ακρίβειας, μεγιστοποιούν την τιμή της κάλυψης, μεγιστοποιώντας, παράλληλα, την πιθανότητα επιβίωσης της πραγματικής κατάστασης.

5.2.3 Πείραμα 2: Εντοπισμός εναλλάσσουσας αποκλίνουσας κατάστασης

Το δεύτερο πείραμα που εκτελέσαμε παρουσιάζει αρκετές ομοιότητες με το πρώτο. Συγκεκριμένα, υπάρχει και πάλι η θεώρηση ότι υπάρχει ένας κακόβουλος χρήστης που επιθυμεί να διαταράξει την ομαλή λειτουργία του δικτύου. Γίνεται, επίσης, πάλι η μοντελοποίηση των χρηστών ως κόμβοι που παράγουν δείγματα σύμφωνα με κάποια κατανομή Poisson. Η κυριότερη διαφορά με το Πείραμα 1 είναι ότι στο προκείμενο πείραμα υποθέτουμε ότι ο κακόβουλος χρήστης μπορεί να αλλάξει ταυτότητα μεταξύ χρονικών παραθύρων. Διαφορετικά, στο φερεμαλισμό του Πειράματος 2 οι κόμβοι αντιστοιχούν, φυσικά, σε διευθύνσεις IP, και όχι σε φυσικούς χρήστες όπως στο Πείραμα 1. Ο λόγος για τον οποίο ένα τέτοιο πείραμα παρουσιάζει πρακτικό ενδιαφέρον είναι επειδή συχνά, κατά την διάρκεια επιθέσεων, οι επιτιθέμενοι χρησιμοποιούν μια τεχνική η οποία ονομάζεται IP spoofing. Αυτή η τεχνική επιτρέπει στους επιτιθέμενους να παρουσιάζουν κάλπικη διεύθυνση IP, και συνεπώς να αποχρύντουν την πραγματική τους ταυτότητα. Σε τέτοιες περιπτώσεις, θα θέλαμε να μπορούμε να εντοπίζουμε τον επιτιθέμενο χρήστη ακόμα και αν αυτό δεν μας προσφέρει άμεσες πληροφορίες για την ταυτότητα του χρήστη, καθώς η ανάλυση των προτύπων των εναλλαγών (ή ακόμα και των ιδίων των διευθύνσεων IP, μέσω ανάλυσης με καμπύλες Hilbert ([57], [1]) μπορούν να μας παρέχουν πληροφορία για τον κακόβουλο χρήστη. Η πειραματική διάταξη παρουσιάζεται σχηματικά στο Σχήμα 5.4.



Σχήμα 5.4: Εντοπισμός εναλλάσσουσας αποκλίνουσας κατάστασης: Ο κακόβουλος χρήστης δύναται να αλλάζει μεταξύ των χρονικών παραθύρων.

Πειραματισμός

Τα επιμέρους πειράματα που εκτελέστηκαν στο ευρύτερο πλαίσιο του Πειράματος 2 παρουσιάζουν εμφανής ομοιότητες με τα επιμέρους πειράματα του Πειράματος 1. Η μοντελοποίηση ακολουθεί πλήρως τα χνάρια του Πειράματος 1: και πάλι, για διαφορετικές τιμές της παραμέτρου χαλάρωσης θ , υπολογίστηκαν οι τιμές των μετρικών ε και ν και καταγράφηκαν οι τιμές κάποιων μεγεθών απαραίτητων για την αξιολόγηση της επίδοσης του αλγορίθμου. Οι τιμές που καταγράφηκαν είναι αυτές του Πειράματος 1, οι οποίες υπενθυμίζονται, συνοπτικά, παρακάτω:

- Η εκάστοτε τιμή της μεταβλητής χαλαρότητας θ , για να είναι δυνατή η ταυτοποίηση των αποτελεσμάτων.
- Η μέση τιμή της μετρικής ε κατά μήκος των πλαισίων.
- Η διαφορά $d\varepsilon$ μεταξύ της μετρικής ε για διαδοχικά ζεύγη τιμών της παραμέτρου χαλάρωσης.
- Η μέση τιμή της μετρικής ν κατά μήκος των πλαισίων.
- Η ακρίβεια εντοπισμού της πραγματικής κατάστασης από τον αλγόριθμο Viterbi με αποκοπή.
- Η κάλυψη της πραγματικής κατάστασης για την εκάστοτε τιμή της παραμέτρου χαλάρωσης.

Θα προχωρήσουμε τώρα στην καταγραφή των παραμέτρων που απαντώνται στα υποπείραματα του Πειράματος 2. Και πάλι, αυτές είναι πανομοιότυπες με εκείνες του Πειράματος 1, τις καταγράφουμε, όμως, για λόγους πληρότητας. Συγκεκριμένα

- Θεωρούμε ότι το πλήθος των χρηστών είναι ίσο με 10, και οι χρήστες κάνουν αιτήματα στο δίκτυο ακολουθώντας κατανομές Poisson.
- Θεωρούμε ότι υπάρχει απουσία πρώιμης γνώσης, και συνεπώς οι *a priori* πιθανότητες είναι μεταξύ τους ίσες, και ισούνται με 0.1, καθώς ο συνολικός αριθμός των χρηστών είναι 10. Οι πιθανότητες μετάβασης τέθηκαν ίσες με αυτές του Πειράματος 1.

- Ο συνολικός αριθμός των χρονικών πλαισίων για τα οποία εκτελέστηκε το εκάστοτε πείραμα τέθηκε ίσος με 10,000.

Θα προχωρήσουμε στην παρουσίαση των αποτελεσμάτων των επιμέρους πειραμάτων που εκτελέστηκαν στο πλαίσιο του Πειράματος 2. Συγκεκριμένα, θα παρουσιάσουμε τρεις πίνακες οι οποίοι αντιστοιχούν σε διαφορετικές επιλογές για την τιμή της παραμέτρου λ_d της κατανομής Poisson του κακόβουλου χρήστη. Οι τιμές που επιλέχθηκαν ταυτίζονται με τις αντίστοιχες του Πειράματος 1, και είναι ίσες με 1, 2, και 3 για τους Πίνακες 5.5, 5.6, και 5.7 αντίστοιχα. Οι τιμές που εξετάσαμε για την παράμετρο χαλάρωσης στα επιμέρους πειράματα ήταν 0, 1, 2, 3, και 4.

Αντίστοιχα με το Πείραμα 1, η ποσότητα λ_d στους τίτλους των πινάκων εκφράζει την παράμετρο της Poisson κατανομής της αποκλίνουσας κατάστασης. Υπενθυμίζουμε ότι, και για αυτή τη συλλογή πειραμάτων, η τιμή της παραμέτρου των Poisson κατανομών των καταστάσεων που αντιστοιχούν σε έγκυρους χρήστες του δικτύου είναι $\lambda = 5$.

Πείραμα εναλλασσόμενης αποκλίνουσας κατάστασης ($\lambda_d = 1$)					
θ	ε	$d\varepsilon$	ν	Ακρίβεια	Κάλυψη
0	13.95484×10^{-6}	-	0	0.7541	0.8285
1	13.87600×10^{-6}	-0.07884×10^{-6}	0.08677	0.7438	0.9418
2	6.02034×10^{-6}	-7.85566×10^{-6}	0.31317	0.7409	0.9847
3	9.23887×10^{-6}	3.21853×10^{-6}	0.56756	0.7494	0.9961
4	10.02658×10^{-6}	0.78771×10^{-6}	0.76713	0.7515	0.9993

Πίνακας 5.5: Τιμές μετρικών ν και ε , ακρίβεια και κάλυψη για διαφορετικές τιμές μεταβλητής χαλάρωσης θ .

Πείραμα εναλλασσόμενης αποκλίνουσας κατάστασης ($\lambda_d = 2$)					
θ	ε	$d\varepsilon$	ν	Ακρίβεια	Κάλυψη
0	12.92497×10^{-6}	-	0	0.5176	0.6115
1	9.31590×10^{-6}	-3.60907×10^{-6}	0.14238	0.5139	0.7889
2	2.45025×10^{-6}	-6.86565×10^{-6}	0.41618	0.4986	0.9054
3	4.89674×10^{-6}	2.44649×10^{-6}	0.63904	0.5006	0.9629
4	1.42705×10^{-6}	-3.46969×10^{-6}	0.75948	0.5086	0.9873

Πίνακας 5.6: Τιμές μετρικών ν και ε , ακρίβεια και κάλυψη για διαφορετικές τιμές μεταβλητής χαλάρωσης θ .

Συμπεράσματα

Σε αυτή την ενότητα θα προβούμε στην ανάλυση των αποτελεσμάτων των επιμέρους πειραμάτων της εναλλασσόμενης αποκλίνουσας κατάστασης. Η ανάλυση θα εκτελεστεί σε δύο στάδια: αρχικά, θα κάνουμε μια ανάλυση των αποτελεσμάτων ως αυτόνομες οντότητες, παρέχοντας καίριες παρατηρήσεις. Έπειτα, θα εξετάσουμε τα αποτελέσματα υπό το πρίσμα της Ενότητας 4.4. Εξετάζοντας τους Πίνακες 5.5, 5.6, και 5.7 μπορούμε να κάνουμε τις εξής παρατηρήσεις για τα αποτελέσματα του Πειράματος 2:

- Είναι εμφανές ότι όταν η διαφορά μεταξύ της αποκλίνουσας κατάστασης και των υπολοίπων είναι μεγάλη, υπάρχει μεγάλο ποσοστό ορθής αναγνώρισης, αντίστοιχα με το

Πείραμα εναλλασσόμενης αποκλίνουσας κατάστασης ($\lambda_d = 3$)					
θ	ε	$d\varepsilon$	ν	Ακρίβεια	Κάλυψη
0	$2.771\ 56 \times 10^{-6}$	-	0	0.3159	0.3861
1	$6.148\ 23 \times 10^{-6}$	$3.376\ 67 \times 10^{-6}$	0.16220	0.3106	0.5870
2	$6.502\ 66 \times 10^{-6}$	$0.354\ 43 \times 10^{-6}$	0.46032	0.3195	0.7578
3	$2.560\ 99 \times 10^{-6}$	$-3.941\ 67 \times 10^{-6}$	0.66693	0.3159	0.8754
4	$5.394\ 81 \times 10^{-6}$	$2.833\ 82 \times 10^{-6}$	0.76863	0.3187	0.9452

Πίνακας 5.7: Τιμές μετρικών ν και ε , ακρίβεια και κάλυψη για διαφορετικές τιμές μεταβλητής χαλάρωσης θ .

Πείραμα 1. Όμως, παρατηρούμε ότι σε γενικές γραμμές τα καταγεγραμμένα ποσοστά ακρίβειας είναι μικρότερα συγκριτικά με αυτά του Πειράματος 1. Κάτι τέτοιο είναι αναμενόμενο, καθώς η διεργασία που καλείται να εκτελέσει ο αλγόριθμος στο Πείραμα 2 είναι σημαντικά πιο σύνθετη από αυτή του Πειράματος 1.

- Παρατηρούμε ότι, αντίστοιχα με το Πείραμα 1, η κάλυψη αυξάνεται με την αύξηση της μεταβλητής χαλάρωσης θ , ακόμα και αν δεν παρατηρείται αντίστοιχη αύξηση στην τιμή της ακρίβειας. Το φαινόμενο αυτό συνάδει πλήρως με τις αντίστοιχες παρατηρήσεις που κάναμε στην ενότητα των συμπερασμάτων του Πειράματος 1.

Κρίνουμε σκόπιμο να κάνουμε μια νύξη στα παρατηρούμενα ποσοστά ορθής αναγνώρισης. Αντίθετα με το Πείραμα 1, όπου η εισαγωγή μη μηδενικής παραμέτρου χαλάρωσης θ συνεπαγόταν άμεση αύξηση της ακρίβειας, παρατηρούμε ότι η εισαγωγή της παραμέτρου χαλάρωσης στο Πείραμα 2 δεν επιφέρει το ίδιο αποτέλεσμα. Κάτι τέτοιο δεν θα έπρεπε να εκπλήσσει, καθώς, όπως αναφέραμε, το Πείραμα 2 καλείται να επιλύσει ένα σημαντικά πιο σύνθετο πρόβλημα απ' ό,τι το Πείραμα 1. Συγκεκριμένα, καθώς υπάρχει δυνατότητα να εναλλάσσεται η αποκλίνουσα κατάσταση, η πρόωμη γνώση που έχει καλλιεργηθεί στα προηγούμενα πλαίσια δεν είναι αξιοποιήσιμη στο τρέχον πλαίσιο, και για την ακρίβεια δρα καταλυτικά στην επιτυχή αναγνώριση, καθώς τείνει να διατηρήσει την πρόβλεψη του κακόβουλου χρήστη στην τρέχουσα κατάσταση. Παρ' όλα αυτά στο προκείμενο μοντέλο θα μπορούσαμε να εισάγουμε διαφορετικά βάρη στους όρους της εξίσωσης Viterbi, έτσι ώστε να δίνεται μεγαλύτερη βαρύτητα στις πιθανότητες παρατήρησης απ' ό,τι στην προσπάθεια παραμονής στην τρέχουσα κατάσταση.

Προχωρώντας στην εξέταση των αποτελεσμάτων των Πινάκων 5.5, 5.6, και 5.7 συνοπλογίζοντας τις παρατηρήσεις της Ενότητας 4.4, έχουμε να σημειώσουμε τα εξής:

- Εν γένει παρατηρούμε ότι οι μεταβολές της ακρίβειας εμπίπτουν στο εύρος της στατιστικής απόκλισης. Τα σχόλια της Ενότητας 4.4 αναφορικά με την τιμή της μετρικής ε δεν απαντώνται στο Πείραμα 2 όπως στο Πείραμα 1. Όπως αναφέραμε αυτό ευθύνεται κυρίως στη δυσκολία της διεργασίας που καλείται να επιλύσει το Πείραμα 2, και πιθανόν να δύναται να μετριαστούν με διαφορετική μοντελοποίηση στην βαρύτητα των όρων της εξίσωσης του Viterbi.
- Στο Πείραμα 2 η μεγάλη διαφορά μεταξύ των τιμών των μετρικών ε συνεπάγεται σημαντική αύξηση στην κάλυψη. Αυτό είναι αναμενόμενο, όπως αναφέραμε και στα συμπεράσματα του Πειράματος 1, και συνάδει με την μέχρι τώρα ανάλυση. Συνεπώς στο προκείμενο πείραμα η χρήση της μετρικής ε αναλύεται σε μια ποσοτικοποίηση του ισοζυγίου μεταξύ ακρίβειας και χρονικής πολυπλοκότητας.

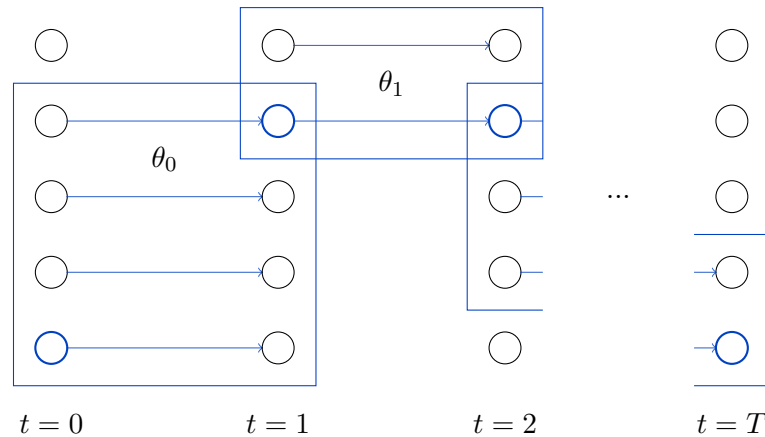
5.2.4 Πείραμα 3: Δυναμική εναλλαγή παραμέτρου χαλάρωσης

Για το τρίτο πείραμα επιλέξαμε να αλλάξουμε λίγο την κατεύθυνση. Συγκεκριμένα, το ευρύτερο πλαίσιο παραμένει ίδιο με τα Πειράματα 1 και 2, δηλαδή θεωρούμε ότι υπάρχει κάποιος κακόβουλος χρήστης που αποσκοπεί στη διαταραχή της ομαλής λειτουργίας του δικτύου, και οι χρήστες μοντελοποιούνται ως κόμβοι που κάνουν αιτήματα στο δίκτυο σύμφωνα με κάποια κατανομή Poisson. Το Πείραμα 3, όμως, αποκλίνει ως προς τον ευρύτερο σκοπό του. Συγκεκριμένα, στο προκείμενο πείραμα αποσκοπούμε στην αυτόματη προσαρμογή της παραμέτρου χαλάρωσης θ . Επιθυμούμε να αλλάζουμε, όποτε κρίνεται αναγκαίο, την τιμή της παραμέτρου χαλάρωσης. Οι λόγοι γι' αυτή τη μοντελοποίηση είναι κυρίως δύο:

- Αφ' ενός, είναι δυνατό διαφορετικές χωρικές περιοχές του δικτύου να έχουν σημαντικά διαφορετικές ιδιότητες μεταξύ τους (για παράδειγμα να υπάρχουν διαφορετικοί βαθμοί συνδεσιμότητας κόμβων σε διαφορετικές χώρες, ανάλογα με τον πληθυσμό της). Συνεπώς, σε μια τέτοια περίπτωση δεν θα θέλαμε να είμαστε δέσμιοι μιας παλαιότερης επιλογής κάποιας παραμέτρου η οποία πλέον δεν επαφίεται στην πραγματικότητα του προβλήματος. Επιπλέον, μια τέτοια θεώρηση μπορεί να επεκταθεί οργανικά στις ιδέες του online learning ([19], [7]). Συγκεκριμένα, στο online learning δίνεται η δυνατότητα το δίκτυο να μεταβάλλεται χρονικά, έτσι ώστε να μοντελοποιήσει τις οργανικές μεταβολές που παρατηρεί ένα δίκτυο. Στο προκείμενο πεδίο εφαρμογής, είναι δυνατό να αλλάζουν οι χρήστες, αλλά σταδιακά. Συνεπώς, η δομή του δικτύου να μεταβάλλεται με μικρά βήματα και αργό ρυθμό. Στα πρώτα στάδια της μεταβολής, μια πρώιμη επιλογή της παραμέτρου χαλάρωσης μπορεί να μοντελοποιεί επαρκώς το ανανεωμένο δίκτυο. Με την πάροδο, όμως, του χρόνου κάτι τέτοιο παύει να ισχύει. Θα θέλαμε, λοιπόν, σε τέτοιες περιπτώσεις, να μεταβάλλουμε αυτόματα την παράμετρο χαλάρωσης, χωρίς να προβούμε ξανά σε ανάλυση του δικτύου.
- Αφ' εταίρου, η παραπάνω μοντελοποίηση επιτρέπει σε μια πιο ευέλικτη αναπαράσταση του κακόβουλου χρήστη. Συγκεκριμένα, ένας κακόβουλος χρήστης μπορεί να κάνει πολλές επιθέσεις, με διαφορετικούς ρυθμούς αιτημάτων κάθε φορά. Η εν λόγω μοντελοποίηση επιτρέπει αυτού του είδους της θεωρήσεις, και είναι σε θέση να αντεπεξέλθει. Επιπλέον, η μοντελοποίηση μπορεί να προσομοιάσει επαρκώς και περιόδους ηρεμίας. Για παράδειγμα, σε περιόδους που δεν πραγματοποιείται κάποια επίθεση υπάρχουν διαφορετικές ανάγκες για την εκτέλεση διαγνωστικών στο δίκτυο, οι οποίες είναι σημαντικά λιγότερο κοστοφόρες αναφορικά με τους διαθέσιμους πόρους. Όταν γίνει εντοπισμός κάποιας επίθεσης, το σύστημα είναι σε θέση, αυτόματα, να επαναπροσδιορίσει την παράμετρο χαλάρωσης, αναθέτοντας περισσότερους πόρους στην εύρεση του κακόβουλου χρήστη και την διατήρηση της ορθής λειτουργίας του συστήματος.

Έχοντας αναλύσει τους λόγους για την παραπάνω μοντελοποίηση, θα αναφερθούμε σύντομα στην πρακτική υλοποίηση του πειράματος. Για την εκτέλεση, επιλέξαμε να ακολουθήσουμε το φορμαλισμό του Πειράματος 1, και συνεπώς να μη δίνουμε τη δυνατότητα στον κακόβουλο χρήστη να αλλάζει καταστάσεις. Ο λόγος που επιλέξαμε τη συγκεκριμένη μοντελοποίηση έναντι αυτής του πειράματος εναλλασόμενης αποκλίνουσας κατάστασης είναι ότι μας ενδιαφέρει απλώς ο εντοπισμός της κατάλληλης παραμέτρου, και όχι η εξέταση των δυνατοτήτων των μετρικών ϵ και ν στην επιλογή της παραμέτρου (με αυτό καταπιάστηκαν τα Πειράματα 1 και 2). Η πειραματική διάταξη με την οποία καταπιάνεται το Πείραμα 3 παρουσιάζεται στο Σχήμα 5.5.

Για την υλοποίηση του Πειράματος 3 αναπτύξαμε τον Αλγόριθμο 3 (Αλγόριθμος Αυτόματης Προσαρμογής Παραμέτρου Χαλάρωσης).



Σχήμα 5.5: Αυτόματη προσαρμογή παραμέτρου χαλάρωσης θ : Η επισημειωμένη κατάσταση σε κάθε χρονικό πλαίσιο υποδηλώνει την κατάσταση μέγιστης πιθανότητας.

Θα αφιερώσουμε λίγο χρόνο σε μερικές παρατηρήσεις αναφορικά με τον Αλγόριθμο 3, και έπειτα θα επεξηγήσουμε, συνοπτικά, τη λειτουργία του. Συγκεκριμένα:

- Για την λειτουργία του αλγορίθμου ζητείται μια αρχική εκτίμηση για τη μεταβλητή χαλάρωσης, θ_0 . Σημειώνεται ότι αυτή η αρχική εκτίμηση δεν επηρεάζει τα αποτελέσματα του αλγορίθμου, καθώς είναι σε θέση, πολύ γρήγορα, να μεταβεί στο κατάλληλο εύρος τιμών για τη μεταβλητή χαλάρωσης.
- Μέχρι να εκτελεστούν τ εποχές χρησιμοποιείται η δοσμένη αρχική εκτίμηση για τη μεταβλητή χαλάρωσης. Κάτι τέτοιο κρίθηκε σκόπιμο, έτσι ώστε οι μέσες τιμές των μετρικών ε και ν να έχουν 'σταθεροποιηθεί' και να είναι ουσιώδης η σύγκριση με τις τιμές του τρέχοντος πλαισίου.
- Αφού περάσουν τ εποχές, και συνεπώς έχει ολοκληρωθεί η διαδικασίας 'μάθησης', συγκρίνεται η τιμή του ε της τρέχουσας επανάληψης με τον μέσο όρο των προηγούμενων τ επαναλήψεων. Εάν, ποσοστιαία, διαφέρουν πάνω από α , τότε επιλέγουμε να αλλάξουμε την μεταβλητή χαλάρωσης.
- Εάν η τιμή του όγκου ν της τρέχουσας επανάληψης είναι μικρότερη του μέσου όρου των προηγούμενων τ επαναλήψεων, τότε θεωρείται ότι το βήμα είναι αντιπαραγωγικό και αυξάνεται η μεταβλητή χαλάρωσης, ποσοστιαία, κατά β . Διαφορετικά, μειώνεται, ποσοστιαία, κατά β .

Διαισθητικά, ο αλγόριθμος διανύει μια περίοδο μάθησης όπου προσπαθεί να αποκτήσει μια εκτίμηση για τις μέσες τιμές των μετρικών ε και ν . Έπειτα, όταν έχει αποκτήσει την παραπάνω εκτίμηση, αξιοποιεί την τιμή της μετρικής ε για να κρίνει εάν είναι αναγκαία η μεταβολή της παραμέτρου χαλάρωσης. Εάν κρίνει ότι υπάρχει σημαντική απόκλιση από την τρέχουσα τιμή της μετρικής ε , τότε αποφασίζει ότι πρέπει να μεταβληθεί η τιμή της παραμέτρου χαλάρωσης. Σε τέτοια περίπτωση, εξετάζει την τρέχουσα τιμή της μετρικής ν και την συγκρίνει με την μέση τιμή των προηγούμενων επαναλήψεων. Εάν η τιμή της μετρικής είναι μεγαλύτερη από τον πρώιμο μέσο όρο, τότε η προκείμενη αλλαγή οδηγεί σε αύξηση του όγκου του χώρου λύσεων, και συνεπώς κρίνεται ότι είναι αναγκαία η αύξηση της παραμέτρου χαλάρωσης.

Algorithm 3 AdaptivePruning($T, n, I, A, \Lambda, \alpha, \beta, \tau, \theta_0$)

```

1:  $t \leftarrow 0$ 
2:  $E, N, L \leftarrow \emptyset$ 
3:  $\theta \leftarrow \theta_0$ 
4: while  $t < T$  do
5:   lattice,  $\varepsilon, \nu \leftarrow \text{viterbi}(n, I, A, \Lambda, \theta)$ 
6:   if  $t \geq \tau$  then
7:     if  $\frac{|\varepsilon - \text{mean}(E)|}{\text{mean}(E)} \geq \alpha$  then
8:       if  $\nu \leq \text{mean}(N)$  then
9:          $\theta \leftarrow (1 + \beta) \times \theta$ 
10:      else
11:         $\theta \leftarrow (1 - \beta) \times \theta$ 
12:      end if
13:    end if
14:  end if
15:  update( $E, N, L$ )
16: end while
17: seq  $\leftarrow$  backtrack( $L$ )
18: return seq,  $L$ 

```

Πειραματισμός

Θα προχωρήσουμε τώρα στον πειραματισμό για τα επιμέρους υποπειράματα του Πειράματος 3. Παρ' όλο που το ευρύτερο πλαίσιο μοντελοποίησης του προκειμένου πειράματος ακολουθεί αυτό του Πειράματος 1, οι τιμές που καταγράφουμε και η μορφοποίηση των πειραμάτων διαφέρει. Συγκεκριμένα, λόγω του ότι η τιμή της παραμέτρου χαλάρωσης θ προσδιορίζεται αυτόματα, τα επιμέρους πειράματα δεν είναι δυνατό να διαχωρίζονται με βάση αυτό το κριτήριο. Έτσι, ως χαρακτηριστικό θεωρούμε πλέον την τιμή της παραμέτρου της κατανομής Poisson του κακόβουλου χρήστη. Γενικότερα, οι τιμές που καταγράφουμε είναι οι εξής:

- Η τιμή της παραμέτρου της κατανομής Poisson λ_d του κακόβουλου χρήστη για κάθε ένα από τα επιμέρους πειράματα.
- Η μέση τιμή, κατά μήκος των χρονικών πλαισίων, των τιμών της παραμέτρου χαλάρωσης θ που υπολογίζει ο Αλγόριθμος 3.
- Η μέση τιμή, κατά μήκος των χρονικών πλαισίων, των καταστάσεων που εξετάστηκαν σε κάθε βήμα από τον Αλγόριθμο 3.
- Η ακρίβεια και η κάλυψη, όπως αυτές ορίστηκαν στα προηγούμενα πειράματα.

Στο πλαίσιο του Πειράματος 3 θα εκτελέσουμε πληθώρα υποπειραμάτων τα οποία, αντίθετα με τα υποπειράματα των Πειραμάτων 1 και 2, έχουν σημαντικές διαφορές μεταξύ τους. Έτσι, για λόγους διευκόλυνσης του αναγνώστη θα παρέχουμε κάποια συμπερασματικά σχόλια μετά το πέρας του κάθε πειράματος, έτσι ώστε τα σχόλια να είναι επίκαιρα και ουσιώδη. Παράλληλα, θα κάνουμε κάποια γενικότερα σχόλια, που αφορούν το σύνολο των της συλλογής πειραμάτων του Πειράματος 3, κατά την ολοκλήρωση της ενότητας.

Για λόγους πληρότητας χρειάζεται να αναφέρουμε τις συγκεκριμένες τιμές που χρησιμοποιήσαμε για τις παραμέτρους του Αλγορίθμου 3 στο σύνολο των υποπειραμάτων του Πειράματος 3, καθώς και τις γενικότερες παραμέτρους των προσομοιώσεων. Συγκεκριμένα:

- Η τιμή της παραμέτρου α αναφορικά με την ποσοστιαία απόκλιση της μέσης τιμής των τιμών της παραμέτρου ε και της τρέχουσας τιμής της τέθηκε ίση με 0.25.
- Η τιμή της παραμέτρου β αναφορικά με την ποσοστιαία αύξηση της παραμέτρου χαλάρωσης θ (βασιζόμενοι στην σχετική τιμή της παραμέτρου ν) τέθηκε ίση με 0.10.
- Η τιμή της παραμέτρου τ αναφορικά με το πλήθος των εποχών εκπαίδευσης τέθηκε ίση με 100.
- Το πλήθος των καταστάσεων τέθηκε ίσο με 10 και οι έγκυροι χρήστες υποθέτουμε ότι έχουν παράμετρο της κατανομής Poisson ίση με $\lambda = 5$. Οι πιθανότητες μετάβασης, όπως και οι *a priori* πιθανότητες τέθηκαν ίσες με αυτές των προηγούμενων πειραμάτων.
- Ο αριθμός των χρονικών πλαισίων για τον οποίον εκτελείται το εκάστοτε πείραμα τέθηκε ίσος με 10,000.

Θα προχωρήσουμε στην παρουσίαση των αποτελεσμάτων των επιμέρους υποπειραμάτων του Πειράματος 3. Συγκεκριμένα, θα παρουσιάσουμε τέσσερις πίνακες που αντιστοιχούν στα τρία διαφορετικά υποπειράματα. Αντίθετα με τα προηγούμενα πειράματα, στο προκείμενο τα επιμέρους μέλη έχουν σημαντικές διαφορές, τις οποίες θα αναλύσουμε σύντομα:

- Το πρώτο υποπείραμα εξετάζει την επίδοση του Αλγορίθμου 3 στον αυτόματο προσδιορισμό της παραμέτρου χαλάρωσης θ , όπου θεωρούμε ότι υπάρχει ένας σταθερός κακόβουλος χρήστης ο οποίος έχει έναν συγκεκριμένο ρυθμό αποστολής αιτημάτων λ_d .
- Το δεύτερο υποπείραμα εξετάζει την επίδοση του Αλγορίθμου 3 όταν θεωρούμε ότι υπάρχει δυνατότητα να αλλάζει ο ρυθμός αποστολής αιτημάτων του κακόβουλου χρήστη. Όπως αναφέραμε και στην εισαγωγή του Πειράματος 3, ένα τέτοιο σενάριο έχει πρακτική σημασία, τόσο αποδίδει ευελιξία στη μοντελοποίηση του συστήματος, αλλά και επειδή επιτρέπει ευελιξία στην μοντελοποίηση του κακόβουλου χρήστη. Για το εν λόγω υποπείραμα παρουσιάζουμε δύο πίνακες, έτσι ώστε να υπάρχει και άμεση σύγκριση με τα αποτελέσματα των προηγούμενων πειραμάτων.
- Το τελευταίο υποπείραμα εξετάζει το ίδιο σενάριο με το προηγούμενο, αλλά έχουμε επιλέξει διαφορετική μοντελοποίηση. Συγκεκριμένα, έχουμε αυξήσει σημαντικά τον αριθμό των καταστάσεων, και επιπλέον έχουμε αλλάξει θεώρηση για τον τρόπο με τον οποίο παράγονται τα αιτήματα, δηλαδή θεωρούμε ότι οι χρήστες ακολουθούν διαφορετική κατανομή για την αποστολή των αιτημάτων τους. Αναλυτικότερες πληροφορίες με το εν λόγω υποπείραμα θα δοθούν στις παραγράφους που προηγούνται του υποπειράματος.

Πείραμα δυναμικής εναλλαγής παραμέτρου χαλάρωσης				
λ_d	mean(θ)	mean(n)	Ακρίβεια	Κάλυψη
1	2.07193	3.4386	0.438	0.7474
2	2.09185	3.0314	0.713	0.9185
3	2.36066	2.309	0.9412	0.9927

Πίνακας 5.8: Μέση τιμή των μεταβλητών χαλάρωσης που υπολογίστηκαν από τον Αλγόριθμο 3, η μέση τιμή του πλήθους των καταστάσεων που εξετάστηκαν, καθώς και η ακρίβεια και η κάλυψη για διαφορετικές τιμές παραμέτρου της αποκλίνουσας κατάστασης.

Ο Πίνακας 5.8 παρουσιάζει τα αποτελέσματα του πρώτου υποπειράματος. Υπενθυμίζεται ότι η παράμετρος της κατανομής Poisson για τους έγκυρους χρήστες είναι $\lambda = 5$. Με

μια πρώτη βλέψη παρατηρούμε κάποια πολύ ευχάριστα και ενθαρρυντικά αποτελέσματα. Συγκεκριμένα, παρατηρούμε ότι, στο συνολικό πλήθος χρονικών πλαισίων, οι επιλογές για την παράμετρο χαλάρωσης θ του Αλγορίθμου 3 οδηγούν σε μεγαλύτερες τιμές ακρίβειας από ότι η χειροκίνητη επιλογή τιμών χαλάρωσης των προηγούμενων πειραμάτων. Η επόμενη ευχάριστη έκπληξη έρχεται όταν εξετάσουμε τη στήλη της κάλυψης. Συγκεκριμένα, παρατηρούμε, ειδικά εάν συγκρίνουμε με τις τιμές των Πινάκων 5.2, 5.3, και 5.4 (οι οποίοι αντιστοιχούν στους πίνακες του Πειράματος 1, στο οποίο κληθήκαμε να αντιμετωπίσουμε το ίδιο πρόβλημα, αλλά με χειροκίνητη επιλογή των τιμών της παραμέτρου χαλάρωσης), ότι οι τιμές κάλυψης είναι μικρές, συγκριτικά με τις μέγιστες τιμές που παρατηρούνταν στους προαναφερθέντες πίνακες. Αυτό είναι ιδιαίτερα ευχάριστο, καθώς συνεπάγεται ότι ο Αλγόριθμος 3 πετυχαίνει μεγάλο ποσοστό ορθής αναγνώρισης, κάνοντας παράλληλα εξοικονόμηση των πόρων του δικτύου, καθώς εξετάζεται σε κάθε πλαίσιο μικρός αριθμός καταστάσεων. Συνεπώς, συμπεραίνουμε ότι οι τιμές της παραμέτρου χαλάρωσης που επιλέγει ο Αλγόριθμος 3 οδηγούν στην καλύτερη δυνατή ακρίβεια, με τη μικρότερη δυνατή σπατάλη πόρων.

Έχοντας ολοκληρώσει τα σχόλια μας για το πρώτο υποπείραμα του Πειράματος 3, θα προχωρήσουμε στην παρουσίαση των αποτελεσμάτων του δεύτερου υποπείραματος. Όπως αναφέραμε και στην παραπάνω εισαγωγή, η ειδοποιός διαφορά με το πρώτο υποπείραμα είναι ότι στο προκείμενο επιτρέπουμε την μεταβολή του ρυθμού αποστολής αιτημάτων του κακόβουλου χρήστη. Επιλέξαμε να μεταβάλλουμε τον ρυθμό αποστολής αιτημάτων του κακόβουλου χρήστη κάθε 1,000 χρονικά πλαίσια. Συγκεκριμένα, ο ρυθμός είτε αυξάνεται κατά μια μονάδα, είτε μειώνεται κατά μια μονάδα, τυχαία. Επιπλέον, για λόγους ορθότητας επιλέξαμε να περιορίσουμε τις δυνατές τιμές που μπορεί να πάρει ο ρυθμός αποστολής αιτημάτων λ_d έτσι ώστε:

$$0 \leq \lambda_d \leq \lambda \quad (5.4)$$

όπου λ είναι ο ρυθμός αποστολής αιτημάτων των έγκυρων χρηστών. Εάν η παράμετρος του κακόβουλου χρήστη τείνει να υπερβεί το επιτρεπτό όριο, ή τείνει να δεχθεί αρνητικές τιμές, τότε την περιορίζουμε αναλόγως. Παρουσιάζονται δύο πίνακες:

- Ο Πίνακας 5.9 παρουσιάζει τα αποτελέσματα για το δεύτερο υποπείραμα, έχοντας υποθέσει ότι οι έγκυροι χρήστες έχουν τιμή παραμέτρου της κατανομής Poisson ίση με $\lambda = 10$, και ο κακόβουλος χρήστης ξεκινάει με τιμή παραμέτρου $\lambda_d = 5$.
- Ο Πίνακας 5.10 παρουσιάζει τα αποτελέσματα για το δεύτερο υποπείραμα, έχοντας υποθέσει ότι οι έγκυροι χρήστες έχουν τιμή παραμέτρου της κατανομής Poisson ίση με $\lambda = 5$, και ο κακόβουλος χρήστης ξεκινάει με τιμή παραμέτρου $\lambda_d = 3$.

Όπως αναφέραμε στο θεωρητικό υπόβαθρο, η κατανομή Poisson έχει διακύμανση ίση με την μέση της τιμή. Αυτό συνεπάγεται ότι, ενώ στο υποπείραμα που παρουσιάζεται στον Πίνακα 5.9 αυξήσαμε την παράμετρο της κατανομής έτσι ώστε να είναι πιο εμφανής οι διαφορές στην τιμή της παραμέτρου χαλάρωσης, η αύξηση των τιμών των κατανομών είχε επίσης αποτέλεσμα να μειωθεί σημαντικά η παρατηρούμενη ακρίβεια, καθώς αυξάνεται σημαντικά η συνολική διακύμανση του υποπείραματος. Κάτι τέτοιο δικαιολογεί το γεγονός ότι, ενώ οι αριθμητικές τιμές της παραμέτρου χαλάρωσης 5 και 6 είναι σημαντικά μικρότερες από την τιμή της παραμέτρου των έγκυρων χρηστών, η ακρίβεια της ορθής αναγνώρισης είναι μικρή. Το παραπάνω φαινόμενο αποτέλεσε και τον κύριο λόγο που επιλέξαμε να εκτελέσουμε και ένα δεύτερο υποπείραμα στο προκείμενο πλαίσιο (αυτό που παρουσιάζεται στον Πίνακα 5.10), έτσι ώστε ο αναγνώστης να μπορεί άμεσα να συγκρίνει τα αποτελέσματα με αυτά του Πειράματος 1, καθώς τα μεγέθη είναι άμεσα συγκρίσιμα. Παράλληλα, επιλέξαμε να διατηρήσουμε και τα αποτελέσματα του Πίνακα 5.9, καθώς μας επέτρεψαν να κάνουμε νύξη στην καταλυτική, εν προκειμένω, επίδραση της διακύμανσης της κατανομής Poisson.

Πείραμα δυναμικής εναλλαγής παραμέτρου χαλάρωσης					
Χιλιάδα	λ_d	mean(θ)	mean(n)	Ακρίβεια	Κάλυψη
1	5	2.07008	2.054	0.696	0.853
2	4	1.55301	1.758	0.845	0.918
3	3	1.05690	1.242	0.916	0.935
4	2	0.70531	1.102	0.985	0.985
5	1	0.67571	1.029	0.996	0.996
6	2	0.69558	1.091	0.978	0.979
7	3	1.07337	1.288	0.932	0.945
8	4	1.72864	1.688	0.845	0.898
9	5	2.08500	2.17	0.708	0.869
10	6	2.37011	2.548	0.532	0.799
Συνολικά	3.5	1.39168	1.597	0.8433	0.9177

Πίνακας 5.9: Μέση τιμή των μεταβλητών χαλάρωσης που υπολογίστηκαν από τον Αλγόριθμο 3, η μέση τιμή του πλήθους των καταστάσεων που εξετάστηκαν, καθώς και η ακρίβεια και η κάλυψη για διαφορετικές τιμές παραμέτρου της αποκλίνουσας κατάστασης.

Πείραμα δυναμικής εναλλαγής παραμέτρου χαλάρωσης					
Χιλιάδα	λ_d	mean(θ)	mean(n)	Ακρίβεια	Κάλυψη
1	3	2.12879	3.619	0.447	0.763
2	2	2.07608	2.922	0.701	0.899
3	3	2.03028	3.572	0.406	0.737
4	2	2.06392	3.029	0.697	0.918
5	3	2.07465	3.446	0.440	0.737
6	2	2.15011	3.023	0.718	0.931
7	1	2.31272	2.328	0.941	0.995
8	2	2.12629	2.954	0.709	0.920
9	1	2.48935	2.324	0.944	0.996
10	0	3.37784	2.148	1.000	1.000
Συνολικά	1.9	2.27205	2.9365	0.6993	0.8896

Πίνακας 5.10: Μέση τιμή των μεταβλητών χαλάρωσης που υπολογίστηκαν από τον Αλγόριθμο 3, η μέση τιμή του πλήθους των καταστάσεων που εξετάστηκαν, καθώς και η ακρίβεια και η κάλυψη για διαφορετικές τιμές παραμέτρου της αποκλίνουσας κατάστασης.

Εξετάζοντας τα αποτελέσματα που παρουσιάζονται στους Πίνακες 5.9 και 5.10 παρατηρούμε ότι, σύμφωνα με τα αποτελέσματα του προηγούμενου υποπειράματος, και πάλι ο Αλγόριθμος 3 προσδιορίζει τις βέλτιστες τιμές για την παράμετρο χαλάρωσης θ . Συγκεκριμένα, οι τιμές που προσδιορίζει, τόσο σε επίπεδο συλλογής εποχών, αλλά και σε συνολικά επίπεδο, για την παράμετρο χαλάρωσης επιτυγχάνουν υψηλά ποσοστά ακρίβειας, χωρίς, όμως, να υπάρχει σπατάλη υπολογιστικών πόρων του συστήματος με ανεξέλεγκτα υψηλά ποσοστά παρατηρούμενης κάλυψης. Κάτι τέτοιο είναι εμφανές και όταν εξετάζουμε τον μέσο αριθμό των καταστάσεων που εξετάστηκαν, ο οποίος επιβεβαιώνει την εξοικονόμηση των πόρων του συστήματος.

Ολοκληρώνοντας τον πειραματισμό οδηγούμαστε και στην ολοκλήρωση του Πειράματος 3, μέσω της παρουσίασης του τελευταίου υποπειράματος της συλλογής. Συγκεκριμένα, επιλέξαμε να παρουσιάσουμε αποτελέσματα από ένα υποπείραμα όπου χρησιμοποιήσαμε σημαντικά

μεγαλύτερο αριθμό καταστάσεων, καθώς και διαφορετική μοντελοποίηση, υπό τη μορφή της διαφορετικής κατανομής για τον ρυθμό αποστολής αιτημάτων. Συγκεκριμένα:

- Ο αριθμός των καταστάσεων τέθηκε ίσος με 100. Αυτή η επιλογή έγινε για να αποδείξουμε, πειραματικά, ότι τα αποτελέσματα γενικεύονται σε δίκτυα αυθαίρετου μεγέθους, όπου παρατηρούνται αντίστοιχα αποτελέσματα με αυτά που προέκυψαν και στα προηγούμενα υποπειράματα.
- Επιλέχθηκε κανονική κατανομή για την μοντελοποίηση του ρυθμού αποστολής αιτημάτων των χρηστών. Συγκεκριμένα, θεωρήσαμε ότι οι έγκυροι χρήστες του δικτύου ακολουθούν κανονική κατανομή με μέση τιμή 10 και τυπική απόκλιση 1. Αντίστοιχα, θεωρήσαμε ότι ο κακόβουλος χρήστης, αρχικά, ακολουθεί κανονική κατανομή μέσης τιμής 8, και τυπικής απόκλισης επίσης 1. Το επιτρεπόμενο εύρος της μέσης τιμής της αποκλίνουσας κατάστασης τέθηκε ίσο με το διάστημα $[5, 15]$, κατ' αναλογία με τον περιορισμό της Εξίσωσης (5.4) του προηγούμενου υποπειράματος.
- Οι αρχικές πιθανότητες τροποποιήθηκαν καταλλήλως έτσι όλες οι καταστάσεις να έχουν την ίδια *a priori* πιθανότητα να είναι κακόβουλες. Επίσης, ο πίνακας μετάβασης αποδίδει 0.05 πιθανότητα να παραμείνει το σύστημα στην ίδια κατάσταση και 0.0095 πιθανότητα να μεταβεί σε οποιαδήποτε από τις άλλες καταστάσεις.

Αξίζει να κάνουμε μια ιδιαίτερη μνεία στους λόγους για τους οποίους επιλέξαμε συγκεκριμένα την κανονική κατανομή για το τελευταίο στάδιο του πειραματισμού, και, κατά συνέπεια, για την ολοκλήρωση της προκείμενης διπλωματικής εργασίας. Συγκεκριμένα, οι λόγοι επιλογής μπορούν να αναλυθούν στους παρακάτω κύριους άξονες:

- Η κανονική κατανομή δύναται να μοντελοποιήσει επαρκώς πολλά προβλήματα από ποικίλλα επιστημονικά πεδία (με μεγαλύτερη συχνότητα εμφάνισης στα πεδία της Βιολογίας, της Φυσικής, και των Οικονομικών), και έτσι η επιβεβαίωση της ορθής λειτουργίας του Αλγορίθμου 3 επιτρέπει την άμεση εφαρμογή του σε πληθώρα προβλημάτων. Επιπλέον, ακόμα και σε προβλήματα που δεν είναι δυνατή η άμεση μοντελοποίηση με τη χρήση μιας κανονικής κατανομής, συχνά οι εκάστοτε κατανομές μπορούν να επεξηγηθούν επαρκώς ως μίξη διάφορων κανονικών κατανομών ([44], [5]).
- Ακόμα και σε τομείς όπου δεν υπάρχει άμεση μοντελοποίηση με κανονικές κατανομές (είτε ως μίξη κανονικών κατανομών, είτε ως μιας μοναδικής κανονικής κατανομής), κάνοντας χρήση του *κεντρικού οριακού θεωρήματος*, η κανονική κατανομή μπορεί να αποβεί χρήσιμη. Πράγματι, σύμφωνα με το θεώρημα το άθροισμα πολλών τυχαίων μεταβλητών συμπεριφέρεται σαν κανονική κατανομή (Θεώρημα 5.1).
- Τέλος, η τυπική απόκλιση των κανονικών κατανομών συμπεριφέρεται διαφορετικά από αυτή των κατανομών Poisson. Συγκεκριμένα, όπως αναφέραμε πριν, στην κατανομή Poisson η διακύμανση είναι πάντα ίση με τη μέση τιμή της κατανομής. Συνεπώς, όταν εξετάζουμε μικρές τιμές της παραμέτρου χαλάρωσης λ , εν γένει θα παρατηρούμε ικανοποιητικά αποτελέσματα, αλλά όταν αυξάνουμε την παράμετρο οι καταστάσεις πολύ γρήγορα γίνονται μη διακρίσιμες. Αντίθετα, στην κανονική κατανομή μπορούμε να επιλέξουμε αυθαίρετα την τιμή που επιθυμούμε για τη διακύμανση, και έτσι να εξετάσουμε την επίδοση του Αλγορίθμου 3 σε διαφορετικές κλίμακες.

Θεώρημα 5.1. Έστω ένα σύνολο τυχαίων μεταβλητών X_1, X_2, \dots, X_n οι οποίες είναι ανεξάρτητες και ακολουθούν την ίδια κατανομή, και ο αριθμός n είναι επαρκώς μεγάλος.

Τότε, η τυχαία μεταβλητή του μέσου όρου τους

$$S_n := \frac{X_1 + X_2 \dots + X_n}{n}$$

ακολουθεί κανονική κατανομή.

Πείραμα δυναμικής εναλλαγής παραμέτρου χαλάρωσης					
Χιλιάδα	λ_d	mean(θ)	mean(n)	Ακρίβεια	Κάλυψη
1	8	1.60332	10.380	0.923	0.970
2	9	0.54447	3.008	0.194	0.273
3	8	1.19683	3.522	0.883	0.925
4	7	1.74523	3.890	0.989	0.989
5	6	0.81491	2.244	0.997	0.997
6	7	2.00325	4.185	0.997	0.997
7	6	0.83713	1.891	0.999	0.999
8	7	2.13624	5.360	0.997	0.998
9	8	1.13196	4.320	0.912	0.947
10	9	0.53398	3.220	0.186	0.287
Συνολικά	7.5	1.26194	4.202	0.8077	0.8382

Πίνακας 5.11: Μέση τιμή των μεταβλητών χαλάρωσης που υπολογίστηκαν από τον Αλγόριθμο 3, η μέση τιμή του πλήθους των καταστάσεων που εξετάστηκαν, καθώς και η ακρίβεια και η κάλυψη για διαφορετικές τιμές παραμέτρου της αποκλίνουσας κατάστασης.

Εξετάζοντας τα αποτελέσματα που παρουσιάζονται στον Πίνακα 5.11 μπορούμε να κάνουμε μερικές σύντομες παρατηρήσεις. Συγκεκριμένα, παρατηρούμε ότι και για την κανονική κατανομή, αλλά και για σημαντικά μεγαλύτερο αριθμό συνολικών καταστάσεων, ο Αλγόριθμος 3 είναι σε θέση να προσαρμόσει κατάλληλα την τιμή της παραμέτρου χαλάρωσης θ έτσι ώστε να μοντελοποιήσει επαρκώς τις χρονικές μεταβολές που παρατηρούνται στο δίκτυο. Πράγματι, παρατηρούμε ότι ο αλγόριθμος οδηγεί συστηματικά σε μεγάλα ποσοστά ακρίβειας, με πλήρως ελεγχόμενα ποσοστά κάλυψης. Οι μόνες περιπτώσεις στις οποίες ο αλγόριθμος αδυνατεί να επιλύσει ικανοποιητικά το πρόβλημα είναι αυτές όπου οι μέσες τιμές των κανονικών κατανομών των έγκυρων χρηστών και του κακόβουλου χρήστη (συνυπολογίζοντας και την επίδραση της διακύμανσης) είναι εξαιρετικά κοντά. Μια περαιτέρω εξέταση του Πίνακα 5.11 επιβεβαιώνει τη δυνατότητα του Αλγορίθμου 3 να εξοικονομεί τους πόρους του συστήματος, όπως υποδηλώνει η στήλη του μέσου αριθμού εξεταζόμενων καταστάσεων σε κάθε χρονικό πλαίσιο.

Συμπεράσματα

Ολοκληρώνοντας το Πείραμα 3, θα αφιερώσουμε την προκείμενη ενότητα σε μερικές γενικές παρατηρήσεις και συμπεράσματα αναφορικά με τον Αλγόριθμο 1. Αντίθετα με τα προηγούμενα πειράματα όπου η αντίστοιχη ενότητα εμπεριείχε αναλυτικά τα αποτελέσματα των πειραμάτων, στο προκείμενο, καθώς αναλύσαμε τα άμεσα συμπεράσματα μετά το πέρας του κάθε υποπειράματος, τα σχόλια θα είναι πιο συνοπτικού χαρακτήρα. Θα θέλαμε να κάνουμε μερικές γενικές παρατηρήσεις, αναφορικά με την επίδοση του Αλγορίθμου 3, αλλά και τη σύνδεση του Πειράματος 3 με το Κεφάλαιο 4:

- Ο Αλγόριθμος 3 οδηγεί σε τιμές της παραμέτρου χαλάρωσης θ οι οποίες είναι σημαντικά

πιο σύνθετες από τις ακέραιες τιμές που επιλέξαμε κατά τη διάρκεια του πρώιμου πειραματισμού. Επιπλέον, παρατηρούμε ότι, ακόμα και για μεγάλες τιμές απόκλισης μεταξύ του ρυθμού αποστολής αιτημάτων των έγκυρων χρηστών και του κακόβουλου χρήστη, ο αλγόριθμος πάλι καταλήγει σε μικρές τιμές για την παράμετρο χαλάρωσης, δηλώνοντας έτσι ρητά την πεποίθησή του για το ισοζύγιο απόδοσης/υπολογιστικής πολυπλοκότητας.

- Το πλήθος των εξεταζόμενων κόμβων που διερευνά ο Αλγόριθμος 3 σε κάθε χρονικό πλαίσιο παραμένει μικρός, ανεξαρτήτως μεγέθους του δικτύου. Αυτή η παρατήρηση διαφέρει από κάποια παρατήρηση αναφορικά με την κάλυψη, καθώς το μέγεθος της κάλυψης εκφράζει μια ποιοτική ανάλυση της χρησιμοποίησης των πόρων του δικτύου. Αντίθετα, ο μέσος όρος των εξεταζόμενων κόμβων αποτελεί ένα ποσοτικό μέγεθος, και το γεγονός ότι ο Αλγόριθμος 3 περιορίζει επαρκώς το πλήθος τους αποτελεί ένα πολύ θεμιτό και αξιόλογο χαρακτηριστικό.
- Ανεξάρτητα τον όγκο του εξεταζόμενου δικτύου, αλλά και ανεξάρτητα της επιλεγμένης μοντελοποίησης για την κατανομή που ακολουθεί ο ρυθμός αποστολής αιτημάτων, παρατηρούμε ότι ο Αλγόριθμος 3 παράγει σθεναρά αποτελέσματα, τα οποία οδηγούν σε υψηλές τιμές ακρίβειας, χωρίς να σπαταλιούνται οι πόροι του δικτύου.

Κεφάλαιο 6

Επίλογος

6.1 Συνεισφορά

Σε αυτό το σημείο, για να ολοκληρώσουμε την ερευνητική εργασία, θα αναλύσουμε επιγραμματικά τις συνεισφορές τις εργασίας. Αυτές αναλύονται στους εξής τομείς:

- Μελέτη και μοντελοποίηση στο πεδίο της τροπικής άλγεβρας των αλγορίθμων προώθησης βαρών, αφαίρεσης ε -μεταβάσεων.
- Μελέτη και μοντελοποίηση στο πεδίο της τροπικής άλγεβρας του αλγορίθμου Viterbi και της παραλλαγής του με αποκοπή. Επιπλέον, επιτελέστηκε γεωμετρική ανάλυση των ιδιοτήτων του αλγορίθμου και αποδόθηκε κάποια ερμηνεία στις ιδιότητες αυτές.
- Ορισμός στατιστικών μεγεθών πάνω στο γεωμετρικό χώρο λύσεων του αλγορίθμου Viterbi με αποκοπή. Συγκεκριμένα, ορίστηκαν οι μετρικές του κανονικοποιημένου όγκου ν και της κανονικοποιημένης εντροπίας ε των πολυτόπων που ορίζονται σε κάθε χρονικό πλαίσιο.
- Συσχέτιση του Cuninghame-Green inverse του πεδίου της τροπικής άλγεβρας με τη διαδικασία της αποκοπής του αλγορίθμου Viterbi.
- Εκτενής πειραματισμός, τόσο με εφαρμογές σε μια δοκιμασία Επεξεργασίας Φυσικής Γλώσσας, αλλά και με ποικιλία προσομοιώσεων στο πεδίο των Τηλεπικοινωνιών.
- Ανάπτυξη αλγορίθμου αυτόματης προσαρμογής της παραμέτρου χαλάρωσης θ του αλγορίθμου Viterbi με αποκοπή.

Σημειώνεται ότι τα αποτελέσματα αναφορικά με την Ενότητα 4.4 παρουσιάστηκαν στο workshop για Signal Processing Advances on Wireless Communications (SPAWC) της IEEE τον Ιούνιο του 2018, και τα αποτελέσματα αναφορικά με την Ενότητα 5.2 θα υποβληθούν για θεώρηση στο conference για Accoustics, Speech, and Signal Processing (ICASSP) της IEEE, που θα λάβει χώρα τον Μάιο του 2019.

6.2 Μελλοντική Έρευνα

Η εργασία σίγουρα δεν έχει εξαντλήσει το αντικείμενο με το οποίο καταπιάστηκε, πόσο μάλλον ολόκληρο τον τομέα. Υπάρχουν πληθώρα κατευθύνσεων προς τις οποίες θα μπορούσε να επεκταθεί η εν λόγω εργασία (και επιθυμούμε να συμβάλουμε σε κάποια από αυτές τις επεκτάσεις). Συγκεκριμένα, οι κύριοι άξονες που θα μπορούσε να επεκταθεί η εργασία είναι οι εξής:

- Θα μπορούσε να γίνει μια μελέτη αναφορικά με τον εκλεπτισμό των μετρικών. Οι μετρικές ν και ε που ορίσαμε πηγάζουν από την γεωμετρία (ως μια έννοια όγκου) και από την θεωρία πληροφορίας (ως μια έννοια της εντροπίας) αντίστοιχα. Όμως, δεν αποτελούν τις μόνες μετρικές του εκάστοτε πεδίου. Ειδικά στον τομέα της γεωμετρίας, υπάρχει πληθώρα μετρικών, όπως είναι η *κυρτότητα*, η *καμπυλότητα*, και άλλες. Επιπλέον, πέραν των γενικών μετρικών που μπορούν να εξαχθούν, υπάρχουν και άλλες μετρικές οι οποίες έχουν εφαρμογή σε συγκεκριμένα πεδία (οι οποίες μπορεί να μην ορίζονται, ή να μην έχουν την απαραίτητη ευρωστότητα σε άλλες εφαρμογές). Παραδείγματα τέτοιων μετρικών είναι το *pitch*, τα *MFCCs*, η συχνότητα *F0* από το πεδίο επεξεργασίας φωνής. Τέτοιες μετρικές μπορεί να μην μπορούν να εφαρμοστούν άμεσα στα πολύτοπα, αλλά η μελέτη, οι ιδέες, και η διαίσθησή τους μπορεί να ωφελήσει, είτε έμμεσα μέσω του εκλεπτισμού των μετρικών, είτε άμεσα μέσω του ορισμού επιπλέον μετρικών.
- Εφαρμογή του *pruning* σε επιπλέον datasets, αποτελεί ίσως την πιο προφανή επέκταση της εργασίας. Τα πειραματικά αποτελέσματα ήταν περιορισμένα, συνεπώς η εφαρμογή των ιδεών σε αναγνωρισμένες βάσεις θα ωρίμαζε τις ιδέες που αναπτύχθηκαν στο πλαίσιο της εργασίας, και θα βοηθούσε περαιτέρω στην οριστικοποίηση των μετρικών, και την επέκταση των εφαρμογών τους. Συγκεκριμένα, όπως αναφέρθηκε και στην Ενότητα 5.2, τα προβλήματα αυτόματης αναγνώρισης φωνής και DNA sequencing αποτελούν οργανικούς στόχους για την εφαρμογή των ιδεών. Επιπλέον, θα είχε ενδιαφέρον η εφαρμογή των ιδεών σε WFSTs τα οποία έχουν κάποια ιδιαίτερη δομή, όπου θα μπορούσε να μελετηθεί, και ίσως να εκμεταλλευτεί η δομή, έτσι ώστε να παραχθούν νέες ιδέες, αποτελέσματα, και διαισθήσεις.
- Ο σχετικός αναγνώστης θα παρατηρήσει ότι υπάρχουν και άλλοι αλγόριθμοι που αναφέρονται στο σύγγραμμα των Hori και Nakamura ([29]) οι οποίοι δεν έχουν μοντελοποιηθεί στα πλαίσια αυτής της εργασίας. Συγκεκριμένα, δεν έχουν μοντελοποιηθεί τρεις αλγόριθμοι οι οποίοι αλλάζουν ριζικά τη δομή των WFST. Αυτοί είναι:
 - ο αλγόριθμος σύνθεσης WFST.
 - ο αλγόριθμος ελαχιστοποίησης WFST.
 - ο αλγόριθμος ντετερμινισμού WFST.

Αυτοί οι αλγόριθμοι παρουσιάζουν κάποιες έμφυτες δυσκολίες και ασυμβατότητες με το πλαίσιο ανάλυσης, και εξ' ου η παράλειψή τους από την ανάλυση. Συγκεκριμένα, η τροπική άλγεβρα, όπως και η γραμμική άλγεβρα, έχουν ως δομικό στοιχείο πίνακες, οι οποίοι αποτελούν στατικές δομές δεδομένων. Αυτό σημαίνει ότι, ενώ τα στοιχεία μπορούν να αλλάζουν συνεχώς, το πλήθος και η συσχετίσεις μεταξύ των στοιχείων παραμένουν σταθερές. Οι αλγόριθμοι αυτοί αλλάζουν ριζικά τη δομή των WFSTs, και την αλλάζουν συνεχώς κατά τη διάρκεια εκτέλεσής τους. Συνεπώς, η ανάγκη τους για δυναμικές μεταβολές στις δομές δεδομένων τους κάνουν ασύμβατους, εν μέρη, με το πλαίσιο ανάλυσης. Παρ' όλα αυτά, η ανάλυσή τους στο πλαίσιο, αν και δύσκολη και πιθανά περιπλεγμένη, πιθανώς να έχει ενδιαφέρον, είτε ως μαθηματική άσκηση, είτε ως καλλιέργεια διαισθητικών οριζόντων, αλλά και για λόγους πληρότητας.

- Το πολύτοπο Newton αποτελεί ένα από τα σημαντικότερα στοιχεία της τροπικής γεωμετρίας. Ορίζεται πάνω σε τροπικά πολυώνυμα, και έχει άμεσες σχέσεις με τον γραμμικό προγραμματισμό, καθώς περιορίζει το χώρο λύσεων. Στην εν λόγω ανάλυση, η εφαρμογή του πολύτοπου Newton θα ήταν καταχρηστική, καθώς τα μεγέθη μας δεν είναι ακριβώς πολυώνυμα, και ακόμα και αν τα θεωρήσουμε ως τέτοια θα ήταν μονώνυμα

(κάτι το οποίο υπονοούμε όταν εκφράζουμε την άποψη ότι οι $n - 1$ -πλευρές του πολυτόπου αποτελούν τα βέλτιστα μονοπάτια). Παρ' όλα αυτά, μελέτη των μεγεθών ως σύνολα βάσεων δεδομένων θα μπορούσε να οδηγήσει σε πιο οργανικό ορισμό τροπικών πολυωνύμων, τα οποία έχουν τη δυνατότητα να επωφεληθούν από το πολύτοπο Newton.

- Ως μια τελευταία επέκταση της τελευταίας εργασίας θα μπορούσαμε να κατονομάσουμε την εξερεύνηση νέων προβλημάτων, αλγορίθμων, και τομέων υπό το πρίσμα της τροπικής γεωμετρίας. Κάτι τέτοιο από μια αρχική σκοπιά μπορεί να φαντάζει εκτός του πλαισίου της εργασίας. Θεωρούμε, όμως, ότι το πλαίσιο της εργασίας είναι η μελέτη της τροπικής γεωμετρίας και η ευλαβική καλλιέργεια της διαίσθησης στον τομέα της επιστήμης των υπολογιστών, και η μελέτη του συγκεκριμένου τομέα των WFSTs δεν ήταν παρά μια εφαρμογή του πλαισίου σε ένα συγκεκριμένο πρόβλημα. Με αυτή την υπόθεση, εφοπλισμένοι με τα συμπεράσματα της εν λόγω ανάλυσης και εφοδιασμένοι με εμπειρία, αποσκοπούμε στην εφαρμογή των ιδεών σε νέους τομείς. Μερικά πιθανά πεδία εφαρμογής είναι:
 - τα Support Vector Machines αποτελούν μοντέλα τα οποία έχουν έμφυτο το μαθηματικό στοιχείο, μέσω τις μοντελοποίησής τους, αλλά επιπλέον είναι εφοδιασμένα με γεωμετρικά στοιχεία. Συνεπώς, μια μελέτη αυτών των μοντέλων μπορεί να οδηγήσει σε τροπικές επεκτάσεις, που πιθανώς να αναγεννήσουν την περιοχή.
 - τα Convolutional Neural Networks αποτελούν μοντέλα τα οποία έχουν γίνει αρκετά δημοφιλή τα τελευταία χρόνια, ειδικά λόγω των επιδόσεών τους στον τομέα της Όρασης Υπολογιστών. Λόγω όμως της έμφυτης δυστροπίας της συνέλιξης, και της αδυναμίας του ανθρώπινου νου να κατανοήσει σε βάθος την διαδικασία, σε συνδυασμό με τη χρήση της σε βαθιές αρχιτεκτονικές κάνει τα CNNs δίκτυα τα οποία, ενώ λειτουργούν με όλο και βελτιούμενους ρυθμούς, έχουν αφανές το στοιχείο τις διαίσθησης. Η μελέτη τους, και η ανάλυσή τους στο πεδίο της τροπικής γεωμετρίας θα μπορούσε να αποβεί καταληκτική: τόσο στο να ξεκλειδωθεί η διαδικασία της συνέλιξης, όσο και των βαθιών νευρωνικών δικτύων, και ίσως η λειτουργία του εγκεφάλου.
 - Ένας πολύ ενδιαφέρον τομέας είναι αυτός της αραιής αναπαράστασης σημάτων. Όντας ένας τομέας ο οποίος έχει ως δομικό στοιχείο την γραμμική άλγεβρα. Συνεπώς, η επέκταση στο πλαίσιο της τροπικής άλγεβρας είναι βατή. Ουσίας θα ήταν ίσως η ανάλυση των αλγορίθμων του τομέα (όπως είναι ο Matching Pursuit), οι οποίοι έχουν έντονο το στοιχείο τις διαίσθησης, καθώς προέρχονται από τη ανάλυση γραμμικών χώρων, υπό του πρίσματος της τροπικής άλγεβρας, σε αναζήτηση νέων δυνατοτήτων.
 - Χρήση της κλασικής γεωμετρίας των γεοδεσικών αποστάσεων και των μη-συμπαγών σωμάτων στο πλαίσιο της τροπικής άλγεβρας. Η εν λόγω γεωμετρία συχνά βαφτίζεται ως υπολογιστική γεωμετρία, και τα κύρια αντικείμενα μελέτης της και οι προσεγγίσεις διαφέρουν σημαντικά από αυτές της τροπικής γεωμετρίας (που ανήκει στον τομέα της αλγεβρικής γεωμετρίας και τις τοπολογικής γεωμετρίας). Όμως, θα ήταν ενδιαφέρον η ανάλυση και η μελέτη της υπό το πρίσμα της τροπικής άλγεβρας, καθώς κάτι τέτοιο έχει δυνατότητα να 'παντρέψει' διαφορετικές πτυχές του τομέα.
 - Τέλος, ίσως η πιο φιλόδοξη επέκταση της εργασίας είναι η εφαρμογή των ιδεών της τροπικής άλγεβρας και της τροπικής γεωμετρίας στον τομέα της επεξεργασίας σημάτων πάνω σε γράφους. Η επεξεργασία σημάτων σε γράφους αποτελεί έναν σχετικά νέο τομέα. Δεδομένης της επιτυχίας των μορφολογικών τελεστών στην

επεξεργασία σημάτων, και της επιτυχίας της τροπικής γεωμετρίας στα γραφικά μοντέλα, μια τέτοια σύνθεση θα αποτελούσε σίγουρα έναν ενδιαφέρον, νέο κλάδο.

Bibliography

- [1] C. Alabiso and I. Weiss, *A Primer on Hilbert Space Theory*. Springer, 2016.
- [2] D. Bertsekas and J. Tsitsiklis, *Introduction to Probability*. Athena Scientific, 2008.
- [3] D. Bertsimas, *Introduction to Linear Optimization*, ser. Athena Scientific Series in Optimization and Neural Computation. Athena Scientific, 1997.
- [4] G. Birkhoff, *Lattice Theory*. American Mathematical Society, 1996.
- [5] C. Bishop, *Pattern Recognition and Machine Learning*, ser. Information Science and Statistics. Springer, 2011.
- [6] A. Bondy and U. Murty, *Graph Theory*. Springer, 2008.
- [7] L. Bottou, «Online Algorithms and Stochastic Approximations», in *Online Learning and Neural Networks*, Cambridge University Press, 1998.
- [8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [9] P. Bremaud, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, ser. Texts in Applied Mathematics. Springer, 2008.
- [10] P. Butkovič, *Max-linear Systems: Theory and Algorithms*. Springer, 2010.
- [11] O. Cappe, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models*, ser. Series in Statistics. Springer, 2007.
- [12] V. Chvatal, *Linear Programming*, ser. Series of Books in Mathematical Sciences. W. H. Freeman, 1983.
- [13] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms*. MIT Press, 2009.
- [14] T. Cover and J. Thomas, *Elements of Information Theory*, ser. Series in Telecommunications and Signal Processing. Wiley-Interscience, 2006.
- [15] R. Cuninghame-Green, *Minimax Algebra*, ser. Lecture Notes in Economics and Mathematical Systems. Springer, 1979.
- [16] B. Davey and H. Priestley, *Introduction to Lattices and Order*. Cambridge University Press, 2002.
- [17] A. Doucet, N. d. Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, ser. Information Science and Statistics. Springer, 2001.
- [18] N. Durrani, H. Sajjad, H. Hoang, and P. Koehn, «Integrating an Unsupervised Transliteration Model into Statistical Machine Learning», in *Proceedings of the International Conference of the European Chapter of the ACL*, ACL, 2014.
- [19] H. Elad, *Introduction to Online Convex Optimization*. 2016.

- [20] G. D. Forney, «The Viterbi Algorithm», *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.
- [21] S. Gaubert and R. D. Katz, «Minimal Half-Spaces and External Representation of Tropical Polyhedra», *Journal of Algebraic Combinatorics*, vol. 33, no. 3, pp. 325–348, 2011.
- [22] C. Godsil and G. F. Royle, *Algebraic Graph Theory*. Springer, 2001.
- [23] G. Gratzner, *Lattice Theory: Foundation*. Birkhauser, 1998.
- [24] W. Greub, *Linear Algebra*, ser. Graduate Texts in Mathematics. Springer, 1981.
- [25] P. Griffiths and J. Harris, *Principles of Algebraic Geometry*. Wiley-Interscience, 1994.
- [26] P. Halmos, *Finite-Dimensional Vector Spaces*, ser. Undergraduate Texts in Mathematics. Springer, 1993.
- [27] R. Hartshorne, *Algebraic Geometry*, ser. Graduate Texts in Mathematics. Springer, 1997.
- [28] A. Holme, *A Royal Road to Algebraic Geometry*. Springer, 2012.
- [29] T. Hori and A. Nakamura, *Speech Recognition Algorithms Using Weighted Finite-State Transducers*, ser. Synthesis Lectures on Speech and Audio Processing. Morgan & Claypool Publishers, 2013.
- [30] M. Hurzeler and H. Kunsch, «Monte Carlo Approximations for General State-Space Models», *Journal of Computational and Graphical Statistics*, vol. 7, no. 2, pp. 175–193, 1998.
- [31] M. Jordan, «Graphical Models», *Statistical Science*, vol. 19, no. 1, pp. 140–155, 2004.
- [32] D. Jurafsky and J. Martin, *Speech and Language Processing*. Prentice Hall, 2009.
- [33] M. Kamien and N. Schwartz, *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management*, ser. Advanced Textbooks in Economics. Elsevier Science, 1991.
- [34] H. Karloff, *Linear Programming*, ser. Modern Birkhauser Classics. Birkhauser, 2008.
- [35] D. Kirk, *Optimal Control Theory: An Introduction*, ser. Dover Books on Electrical Engineering. Dover Publications, 2004.
- [36] K. Knight and J. Graehl, «Machine Transliteration», *Journal of Computational Linguistics*, vol. 24, no. 4, pp. 599–612, 1998.
- [37] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, ser. Adaptive Computation and Machine Learning. MIT Press, 2009.
- [38] D. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, ser. International Series in Operations Research Management Science. Springer, 2016.
- [39] D. Maclagan and B. Sturmfels, *Introduction to Tropical Geometry*. American Mathematical Society, 2015.
- [40] P. Maragos, «Dynamical Systems on Weighted Lattices: General Theory», *Mathematics of Control, Signals, and Systems*, vol. 29, no. 21, pp. 1–33, 2017.
- [41] M. Mohri, «Weighted Automata Algorithms», in *Handbook of Weighted Automata*, Springer, 2009, pp. 213–254.
- [42] M. Mohri, «Minimization Algorithms for Sequential Transducers», *Theoretical Computer Science*, vol. 234, pp. 177–201, 2000.

-
- [43] M. Mohri, F. Pereira, and M. Riley, «Weighted Finite-State Transducers in Speech Recognition», *Computer Speech & Language*, vol. 16, no. 1, pp. 69–88, 2002.
- [44] K. Murphy, *Machine Learning: A Probabilistic Perspective*, ser. Adaptive Computation and Machine Learning. MIT Press, 2012.
- [45] C. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, ser. Dover Books on Computer Science. Dover Publications, 1998.
- [46] J. Pitman, *Probability*, ser. Springer Texts in Statistics. Springer, 1999.
- [47] L. R. Rabiner, «A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition», *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [48] C. Shannon and W. Weaver, *The Mathematical Theory of Communication*. University of Illinois Press, 1998.
- [49] K. Smith, L. Kahanpaa, P. Kekalainen, and W. Traves, *An Invitation to Algebraic Geometry*. Springer, 2004.
- [50] J. Stone, *Information Theory: A Tutorial Introduction*. Sebtel Press, 2015.
- [51] G. Strang, *Linear Algebra and Its Applications*. Cengage Learning, 2006.
- [52] A. Viterbi, «Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm», *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.
- [53] M. Wainwright and M. Jordan, *Graphical Models, Exponential Families, and Variational Inference*, ser. Foundations and Trends in Machine Learning. Now Publishers Inc, 2008.
- [54] C. Wang, K. Cohen, and Q. Zhao, «Active Hypothesis Testing on a Tree: Anomaly Detection Under Hierarchical Observations», in *Proceedings of the International Symposium on Information Theory*, IEEE, 2017.
- [55] C. Wang, Q. Zhao, and K. Cohen, «Dynamic Search on a Tree with Information-Directed Random Walk», in *Proceedings of the Workshop on Signal Processing Advances in Wireless Communications*, IEEE, 2018.
- [56] D. West, *Introduction to Graph Theory*. Pearson, 2000.
- [57] N. Young, *An Introduction to Hilbert Space*. Cambridge University Press, 1988.
- [58] G. M. Ziegler, *Lectures on Polytopes*. Springer, 2012.

