ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Πτυχιακή Εργασία

Τίτλος Πτυχιακής Εργασίας	«Ευφυή Συστήματα Συστάσεων: Τεχνικές και Υλοποίηση» «Intelligent Recommendation Systems: Techniques and Implementation»
Ονοματεπώνυμο Φοιτητή	Χρήστος Βεργής
Πατρώνυμο	Αθανάσιος
Αριθμός Μητρώου	Π20029
Επιβλέπων	Κωνσταντίνος Λιαγκούρας, Επίκουρος Καθηγητής

Ημερομηνία Παράδοσης Σεπτέμβριος 2026

Copyright ©

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στον επιβλέποντα καθηγητή μου, κ. Κωνσταντίνο Λιαγκούρα, για την καθοδήγηση, την υποστήριξη και τις εποικοδομητικές παρατηρήσεις του καθ' όλη τη διάρκεια εκπόνησης της παρούσας πτυχιακής εργασίας.

Ένα ιδιαίτερο ευχαριστώ οφείλω στην οικογένειά μου, για την αμέριστη συμπαράσταση, την υπομονή και την ενθάρρυνση που μου προσέφεραν, όχι μόνο κατά τη διάρκεια εκπόνησης αυτής της εργασίας, αλλά σε όλη τη διάρκεια των σπουδών μου.

Τέλος, θα ήθελα να ευχαριστήσω τους φίλους μου για τη συνεχή στήριξη και την κατανόηση που έδειξαν, ιδιαίτερα κατά τις περιόδους εντατικής ασχολίας με την παρούσα εργασία.

Περίληψη

Η ραγδαία αύξηση του όγκου της διαθέσιμης ψηφιακής πληροφορίας έχει αναδείξει τα συστήματα συστάσεων (Recommendation Systems) σε έναν από τους πιο κρίσιμους τομείς εφαρμογής της τεχνητής νοημοσύνης, με εφαρμογές που εκτείνονται από το ηλεκτρονικό εμπόριο και τις υπηρεσίες ροής περιεχομένου έως την εκπαίδευση, την υγεία και τη γεωργία. Η παρούσα πτυχιακή εργασία πραγματοποιεί συστηματική βιβλιογραφική ανασκόπηση εκατό πρόσφατων επιστημονικών δημοσιεύσεων, κυρίως της περιόδου 2020-2026, με στόχο να χαρτογραφήσει συγκριτικά την εξέλιξη, τις αρχιτεκτονικές και τις προκλήσεις του πεδίου.

Η εργασία ξεκινά από τις κλασικές μεθόδους συνεργατικής διήθησης και παραγοντοποίησης πίνακα, προχωρά στις αρχιτεκτονικές βαθιάς μάθησης (όπως οι αυτόματοι κωδικοποιητές, το Neural Collaborative Filtering και οι μηχανισμοί Self-Attention) και εξετάζει εκτενώς τα Γραφικά Νευρωνικά Δίκτυα, τα οποία αναπαριστούν τις αλληλεπιδράσεις χρήστη-αντικειμένου ως γράφο, επιτρέποντας πιο εκφραστική μοντελοποίηση πολύπλοκων σχέσεων.

Ιδιαίτερη έμφαση δίνεται σε ειδικές κατηγορίες συστημάτων (context-aware, multi-behavior, ερμηνεύσιμα, με επίγνωση δικαιοσύνης), καθώς και στους κυριότερους τομείς εφαρμογής, από το παραδοσιακό ηλεκτρονικό εμπόριο έως αναδυόμενα πεδία όπως η υγεία, η γεωργία και η κυβερνοασφάλεια. Η εργασία αναλύει επίσης κριτικά τις κεντρικές προκλήσεις του πεδίου (αραιότητα δεδομένων (data sparsity), ψυχρή εκκίνηση, μεροληψία δημοτικότητας, ζητήματα ιδιωτικότητας και ανθεκτικότητας σε ανταγωνιστικές επιθέσεις) και παρουσιάζει συγκριτική αξιολόγηση των κυριότερων αρχιτεκτονικών βάσει αναφερόμενων αποτελεσμάτων της βιβλιογραφίας.

Τα ευρήματα αναδεικνύουν σαφή τάση μετάβασης από τις κλασικές προς τις αρχιτεκτονικές βαθιάς μάθησης και γράφων, παράλληλα με αυξανόμενο ενδιαφέρον για ζητήματα δικαιοσύνης, ερμηνευσιμότητας και υπεύθυνης τεχνητής νοημοσύνης, θέτοντας τις βάσεις για μελλοντική έρευνα σε κατευθύνσεις όπως η ενσωμάτωση Μεγάλων Γλωσσικών Μοντέλων και οι αρχιτεκτονικές εξατομίκευσης on-device.

Θεματική Περιοχή

Συστήματα Συστάσεων, Τεχνητή Νοημοσύνη, Μηχανική Μάθηση

Λέξεις Κλειδιά

Συστήματα Συστάσεων, Συνεργατική Διήθηση, Βαθιά Μάθηση, Γραφικά Νευρωνικά Δίκτυα, Παραγοντοποίηση Πίνακα, Μηχανισμοί Προσοχής, Δικαιοσύνη Αλγορίθμων

Abstract

The rapid growth in the volume of available digital information has established Recommendation Systems as one of the most critical application domains of artificial intelligence, with use cases spanning e-commerce, streaming services, education, healthcare, and agriculture. This thesis conducts a systematic literature review of one hundred recent scientific publications, primarily from the 2020-2026 period, aiming to holistically map the evolution, architectures, and challenges of the field.

The work begins with classical collaborative filtering and matrix factorization methods, progresses to deep learning architectures (such as autoencoders, Neural Collaborative Filtering, and Self-Attention mechanisms) and extensively examines Graph Neural Networks, which represent user-item interactions as a graph, enabling more expressive modeling of complex relationships.

Particular emphasis is placed on specialized categories of recommender systems (context-aware, multi-behavior, explainable, fairness-aware), as well as on the main application domains, ranging from traditional e-commerce to emerging fields such as healthcare, agriculture, and cybersecurity. The thesis also critically analyzes the central challenges of the field (data sparsity, cold start, popularity bias, privacy concerns, and robustness against adversarial attacks) and presents a comparative evaluation of the main architectures based on results reported in the literature.

The findings reveal a clear trend of transition from classical approaches toward deep learning and graph-based architectures, alongside growing interest in fairness, interpretability, and responsible artificial intelligence, laying the groundwork for future research directions such as the integration of Large Language Models and on-device personalization architectures.

Subject Area

Recommendation Systems, Artificial Intelligence, Machine Learning

Keywords

Recommendation Systems, Collaborative Filtering, Deep Learning, Graph Neural Networks, Matrix Factorization, Attention Mechanisms, Algorithmic Fairness

Πίνακας Περιεχομένων

Κατάλογος Διαγραμμάτων και Σχημάτων

Κατάλογος Πινάκων

Κατάλογος Αλγορίθμων

Εισαγωγή

Η ραγδαία ανάπτυξη του διαδικτύου και η εκθετική αύξηση της διαθέσιμης ψηφιακής πληροφορίας έχουν δημιουργήσει ένα φαινόμενο που στη βιβλιογραφία αναφέρεται ως πληροφοριακή υπερφόρτωση (information overload). Σε έναν κόσμο όπου οι χρήστες καλούνται να επιλέξουν ανάμεσα σε εκατομμύρια προϊόντα, ταινίες, άρθρα ή μουσικά κομμάτια, η δυνατότητα αυτόματης και εξατομικευμένης πρότασης κατάλληλου περιεχομένου έχει καταστεί αναγκαιότητα και όχι πολυτέλεια.

Τα συστήματα συστάσεων (recommendation systems) αποτελούν σήμερα έναν από τους πιο ευρέως χρησιμοποιούμενους τύπους εφαρμογών τεχνητής νοημοσύνης, παρόντα στις πλατφόρμες ροής βίντεο (YouTube, Netflix), στα ηλεκτρονικά καταστήματα (Amazon, eBay), στις υπηρεσίες μουσικής (Spotify), αλλά και σε τομείς υγείας, εκπαίδευσης και τουρισμού. Σκοπός τους είναι να εκτιμήσουν τις προτιμήσεις του χρήστη και να παρουσιάσουν αντικείμενα που πιθανώς θα τον ενδιέφεραν, βασιζόμενα σε ιστορικά δεδομένα αλληλεπίδρασης χρήστη-αντικειμένου.

Οι πρώτες προσεγγίσεις στο πεδίο στηρίχθηκαν κυρίως στις τεχνικές collaborative filtering και content-based filtering, οι οποίες, αν και αποτελεσματικές για απλά σενάρια, αντιμετωπίζουν σημαντικά προβλήματα σε πραγματικές εφαρμογές μεγάλης κλίμακας, όπως η αραιότητα δεδομένων (data sparsity), το πρόβλημα του νέου χρήστη (cold start) και η αδυναμία σύλληψης σύνθετων μη γραμμικών σχέσεων. Η είσοδος των τεχνικών βαθιάς μάθησης (deep learning) στο πεδίο άλλαξε ριζικά τα δεδομένα, καθώς τα νευρωνικά δίκτυα κατάφεραν να μοντελοποιήσουν πολύπλοκες αλληλεπιδράσεις μεταξύ χρηστών και αντικειμένων με πολύ μεγαλύτερη ακρίβεια.

Παράλληλα, η εμφάνιση των Γραφικών Νευρωνικών Δικτύων (Graph Neural Networks, GNNs) άνοιξε νέους δρόμους, επιτρέποντας τη φυσική αναπαράσταση των σχέσεων χρηστών-αντικειμένων ως γράφους και αξιοποιώντας τη δομή του γράφου για τη βελτίωση των συστάσεων. Σύγχρονες αρχιτεκτονικές όπως τα Transformers, αρχικά σχεδιασμένες για την επεξεργασία φυσικής γλώσσας, εφαρμόζονται πλέον επιτυχώς και στο πεδίο των συστημάτων συστάσεων, παρέχοντας εξελιγμένους μηχανισμούς προσοχής (attention mechanisms) που βελτιώνουν σημαντικά την ποιότητα των αποτελεσμάτων.

1.1 Σκοπός και αντικείμενο της εργασίας

Η παρούσα πτυχιακή εργασία έχει ως στόχο τη συνολική αποτύπωση και κριτική ανάλυση των ευφυών τεχνικών που χρησιμοποιούνται για την ανάπτυξη σύγχρονων συστημάτων συστάσεων. Συγκεκριμένα, εξετάζεται η εξέλιξη των μεθόδων από τις κλασικές προσεγγίσεις μέχρι τα πιο πρόσφατα μοντέλα βαθιάς μάθησης και γραφικών νευρωνικών δικτύων, με έμφαση στις ευφυείς τεχνικές που επιτρέπουν τη βελτίωση της ακρίβειας, της κλιμακωσιμότητας (scalability) και της ερμηνευσιμότητας των συστάσεων.

Ειδικότερα, η εργασία επικεντρώνεται στα εξής ερευνητικά ζητήματα: ποιες είναι οι κυριότερες κατηγορίες ευφυών τεχνικών που εφαρμόζονται στα συστήματα συστάσεων, πώς εξελίχθηκαν στο χρόνο, ποιες είναι οι δυνατότητες και οι αδυναμίες κάθε προσέγγισης, και ποιες είναι οι τάσεις και οι προκλήσεις της σύγχρονης έρευνας στο πεδίο. Η εργασία δεν εστιάζει στην υλοποίηση ενός συγκεκριμένου συστήματος, αλλά παρέχει μια συνθετική και τεκμηριωμένη βιβλιογραφική επισκόπηση που αξιοποιεί ένα εκτεταμένο σύνολο επιστημονικών δημοσιεύσεων από περιοδικά και συνέδρια υψηλής απήχησης.

1.2 Μεθοδολογική προσέγγιση

Για την εκπόνηση της εργασίας ακολουθήθηκε συστηματική μεθοδολογία βιβλιογραφικής ανασκόπησης. Η αναζήτηση πηγών πραγματοποιήθηκε σε έγκυρες επιστημονικές βάσεις δεδομένων, όπως το IEEE Xplore, το ScienceDirect, το SpringerLink και το Google Scholar, με χρήση κατάλληλων λέξεων-κλειδιών όπως "recommendation systems", "deep learning", "graph neural networks", "collaborative filtering", "sequential recommendation" και συνδυασμούς αυτών. Τα άρθρα που συμπεριελήφθησαν δημοσιεύθηκαν κυρίως κατά την περίοδο 2020–2026, με έμφαση στις πιο πρόσφατες εξελίξεις.

Συνολικά αναλύθηκαν εκατό (100) επιστημονικά άρθρα, τα οποία καλύπτουν ένα ευρύ φάσμα θεμάτων: από τις κλασικές μεθόδους collaborative filtering και matrix factorization, μέχρι αρχιτεκτονικές βαθιάς μάθησης (νευρωνικά δίκτυα, RNN, CNN, Transformers, GNN), καθώς και εξειδικευμένες κατηγορίες συστημάτων, όπως τα context-aware, sequential, trust-based και

explainable recommendation systems. Η επιλογή των άρθρων διασφάλισε τόσο την ευρύτητα κάλυψης του πεδίου όσο και την επιστημονική αξιοπιστία των πηγών.

1.2.1 Κατανομή papers ανά κατηγορία

Από τα 100 τελικά επιλεγμένα papers, 31 αφορούν εφαρμογές ηλεκτρονικού εμπορίου, 22 τον κινηματογράφο και ψυχαγωγία, 15 την εκπαίδευση, 11 τον τουρισμό, 8 τα τρόφιμα, 7 τα κοινωνικά δίκτυα και 6 άλλους τομείς. Ως προς την αρχιτεκτονική, 28 papers χρησιμοποιούν GNN-based προσεγγίσεις, 22 Transformer-based, 15 RNN/LSTM, 12 CNN-based, 11 MF/NCF, 8 υβριδικά μοντέλα και 4 άλλες αρχιτεκτονικές. Η χρονολογική κατανομή δείχνει σαφή αύξηση των GNN και Transformer papers μετά το 2021, συμπίπτοντας με την ευρεία υιοθέτηση του LightGCN και του SASRec ως βασικών αρχιτεκτονικών αναφοράς.

1.3 Διάρθρωση της εργασίας

Η εργασία οργανώνεται σε δέκα κεφάλαια, τα οποία δεν έχουν όλα το ίδιο βάρος: τα Κεφάλαια 2 και 3 θέτουν το θεωρητικό και ιστορικό υπόβαθρο, ενώ τα Κεφάλαια 4 και 5 αποτελούν, κατά την κρίση μου, τον πυρήνα της εργασίας, καθώς εκεί εξετάζονται αναλυτικά οι αρχιτεκτονικές βαθιάς μάθησης και γράφων που καθορίζουν τη σημερινή αιχμή του πεδίου.

Τα Κεφάλαια 6 έως 8 μετατοπίζουν σκόπιμα την οπτική από το «πώς λειτουργεί» ένα μοντέλο στο «πού εφαρμόζεται και τι πάει στραβά» — μετάβαση που θεωρώ απαραίτητη, καθώς αρκετές ανασκοπήσεις του πεδίου σταματούν στην τεχνική περιγραφή χωρίς να συζητούν επαρκώς περιορισμούς και ανοιχτά ζητήματα. Το ένατο κεφάλαιο επιχειρεί να συγκεντρώσει τα σκόρπια ποσοτικά ευρήματα σε συγκρίσιμη μορφή, κάτι που αποδείχθηκε δυσκολότερο απ' όσο είχα αρχικά υποθέσει λόγω της ανομοιογένειας των πρωτοκόλλων αξιολόγησης μεταξύ των άρθρων (βλ. Ενότητα 9.9), και η εργασία ολοκληρώνεται με το δέκατο κεφάλαιο.

1.4 Οικονομική και κοινωνική σημασία των συστημάτων συστάσεων

Στα δεδομένα που αναφέρει η βιβλιογραφία, η ενσωμάτωση συστημάτων συστάσεων σε πλατφόρμες ηλεκτρονικού εμπορίου συνδέεται συστηματικά με άνοδο του conversion rate και της μέσης αξίας παραγγελίας, χωρίς ωστόσο τα επιμέρους άρθρα να συγκλίνουν σε συγκεκριμένο εύρος τιμών — κάτι αναμενόμενο, καθώς τα ποσοστά εξαρτώνται έντονα από τον κλάδο, το μέγεθος του καταλόγου προϊόντων και τη μέθοδο μέτρησης, οπότε θα ήταν παραπλανητικό να παρατεθεί εδώ ένας ενιαίος αριθμός σαν να επρόκειτο για σταθερά. Σε υπηρεσίες ροής περιεχομένου, το ίδιο μοτίβο επαναλαμβάνεται για τη διατήρηση της προσοχής του χρήστη και τη μείωση της εγκατάλειψης της υπηρεσίας (churn). Η ανάλυση των Peng και Li (2026) επιβεβαιώνει ότι η ενσωμάτωση ευφυών τεχνικών σύστασης αποτελεί πλέον βασικό παράγοντα ανταγωνιστικού πλεονεκτήματος σε πολυάριθμους κλάδους της ψηφιακής οικονομίας, από τα ψηφιακά μέσα έως την εκπαίδευση και τον τουρισμό.

Παράλληλα με την οικονομική διάσταση, τα συστήματα συστάσεων έχουν σημαντικό κοινωνικό αντίκτυπο. Από τη μία πλευρά, συμβάλλουν στη μείωση της πληροφοριακής υπερφόρτωσης, διευκολύνοντας τον χρήστη να εντοπίσει σχετικό περιεχόμενο μέσα σε ένα συνεχώς διογκούμενο σύνολο επιλογών. Από την άλλη, η άκριτη βελτιστοποίηση μετρικών αφοσίωσης (engagement) μπορεί να οδηγήσει σε ανεπιθύμητα φαινόμενα, όπως ο εγκλωβισμός του χρήστη σε «θαλάμους ηχούς» περιεχομένου ή η ενίσχυση ακραίου περιεχομένου. Το ζήτημα αυτό αναλύεται εκτενέστερα στο Κεφάλαιο 8, όπου εξετάζονται οι προκλήσεις δικαιοσύνης, διαφάνειας και κοινωνικής ευθύνης που συνοδεύουν τη σχεδίαση σύγχρονων συστημάτων συστάσεων.

1.5 Δεοντολογικό πλαίσιο και αναπαραγωγιμότητα της έρευνας

Η παρούσα εργασία βασίζεται αποκλειστικά σε δημόσια διαθέσιμη επιστημονική βιβλιογραφία, χωρίς συλλογή πρωτογενών δεδομένων από πραγματικούς χρήστες, και συνεπώς δεν εγείρει άμεσα ζητήματα προστασίας προσωπικών δεδομένων κατά την εκπόνησή της. Ωστόσο, η συστηματική ανασκόπηση εκατό επιστημονικών άρθρων απαιτεί προσεκτική και

διαφανή τεκμηρίωση της μεθοδολογίας αναζήτησης, επιλογής και ανάλυσης πηγών, ώστε τα συμπεράσματα της εργασίας να μπορούν να ελεγχθούν και να αναπαραχθούν από τρίτους ερευνητές, σύμφωνα με τις αρχές της ανοιχτής και αναπαραγώγιμης επιστήμης.

Η αναπαραγωγιμότητα αποτελεί ιδιαίτερα κρίσιμο ζήτημα στο πεδίο των συστημάτων συστάσεων, καθώς πολλές δημοσιεύσεις δεν συνοδεύονται από δημόσια διαθέσιμο κώδικα ή πλήρη περιγραφή των υπερπαραμέτρων εκπαίδευσης, γεγονός που δυσχεραίνει την ανεξάρτητη επαλήθευση των αναφερόμενων αποτελεσμάτων. Για τον λόγο αυτό, στην παρούσα εργασία γίνεται σαφής διάκριση μεταξύ ευρημάτων που βασίζονται σε άμεσα αναφερόμενα αριθμητικά αποτελέσματα της βιβλιογραφίας και ενδεικτικών παραδειγμάτων ή απλουστευμένων υπολογισμών που κατασκευάστηκαν για εκπαιδευτικούς σκοπούς εντός της εργασίας, όπως το αριθμητικό παράδειγμα της Ενότητας 2.8.

1.6 Περιορισμοί της παρούσας εργασίας

Η παρούσα εργασία, ως βιβλιογραφική ανασκόπηση, υπόκειται σε ορισμένους εγγενείς περιορισμούς που πρέπει να ληφθούν υπόψη κατά την ερμηνεία των ευρημάτων της. Πρώτον, η επιλογή των εκατό άρθρων, παρότι ακολούθησε συστηματική μεθοδολογία αναζήτησης (Ενότητα 1.2), δεν μπορεί να θεωρηθεί πλήρως εξαντλητική σε σχέση με το σύνολο της διεθνούς βιβλιογραφίας, η οποία εξελίσσεται με ταχύτατο ρυθμό. Δεύτερον, η έμφαση σε άρθρα της περιόδου 2020-2026 ενδέχεται να υποαντιπροσωπεύει θεμελιωδώς σημαντικές, παλαιότερες συμβολές στο πεδίο, οι οποίες ωστόσο εξακολουθούν να αποτελούν τη θεωρητική βάση πολλών σύγχρονων τεχνικών.

Τρίτον, η εργασία δεν περιλαμβάνει πρωτότυπη πειραματική υλοποίηση ή αξιολόγηση μοντέλων σε πραγματικά δεδομένα, αλλά βασίζεται στη σύνθεση και κριτική ανάλυση αναφερόμενων αποτελεσμάτων από τρίτες πηγές. Συνεπώς, οι ποσοτικές συγκρίσεις που παρουσιάζονται στο Κεφάλαιο 9 και στο Παράρτημα Β θα πρέπει να ερμηνεύονται ως ενδεικτικές τάσεις της βιβλιογραφίας και όχι ως αποτελέσματα ελεγχόμενων, συγκριτικών πειραμάτων υπό ταυτόσημες συνθήκες, περιορισμό που συζητείται περαιτέρω στην Ενότητα 9.9.

1.7 Σχετικές βιβλιογραφικές ανασκοπήσεις και διαφοροποίηση της εργασίας

Η παρούσα εργασία δεν αποτελεί την πρώτη βιβλιογραφική ανασκόπηση στο πεδίο των συστημάτων συστάσεων· αντίθετα, εντάσσεται σε μια ευρύτερη παράδοση συστηματικών ανασκοπήσεων που έχουν επιχειρήσει να χαρτογραφήσουν την εξέλιξη του πεδίου από διαφορετικές οπτικές γωνίες. Ορισμένες πρόσφατες ανασκοπήσεις εστιάζουν αποκλειστικά σε αρχιτεκτονικές βαθιάς μάθησης, άλλες υιοθετούν αυστηρή μεθοδολογία τύπου PRISMA με έμφαση σε πολυκριτηριακά συστήματα, ενώ άλλες εξετάζουν αποκλειστικά γραφικά νευρωνικά δίκτυα.

Η διαφοροποίηση της παρούσας εργασίας έγκειται στην προσπάθεια ευρύτερης κάλυψης του πεδίου, καλύπτοντας ταυτόχρονα το πλήρες φάσμα αρχιτεκτονικών (από κλασικές μεθόδους έως GNN και Transformer), πολλαπλούς τομείς εφαρμογής, και κριτική ανάλυση ανοιχτών προκλήσεων, αντί να εξειδικεύεται σε μία μόνο αρχιτεκτονική οικογένεια ή έναν μόνο τομέα εφαρμογής. Θεωρώ ότι οι ανασκοπήσεις τύπου PRISMA, παρότι μεθοδολογικά αυστηρότερες, τείνουν να θυσιάζουν αυτού του είδους το εύρος για χάρη της αναπαραγωγιμότητας· η δική μου επιλογή να μην ακολουθήσω αυστηρό πρωτόκολλο ένταξης/αποκλεισμού τύπου PRISMA έγινε συνειδητά, επειδή στόχος εδώ δεν ήταν η μετα-ανάλυση αλλά η ανάδειξη διασυνδέσεων μεταξύ φαινομενικά ανεξάρτητων υποπεδίων (όπως η σύνδεση μεταξύ over-smoothing στα GNN (Ενότητα 5.12) και ανθεκτικότητας σε ανταγωνιστικές επιθέσεις (Ενότητα 8.7.1)), οι οποίες συχνά παραβλέπονται σε πιο εξειδικευμένες ανασκοπήσεις.

Αναγνωρίζω πάντως ότι αυτή η επιλογή έχει κόστος διπλής φύσης: αφενός συνεπάγεται μικρότερο βάθος ανάλυσης σε κάθε επιμέρους υποπεδίο σε σχέση με μια στενότερα εστιασμένη ανασκόπηση (περιορισμός που συζητήθηκε ήδη στην Ενότητα 1.6), αφετέρου η απουσία αυστηρών κριτηρίων ένταξης σημαίνει ότι η επιλογή των εκατό άρθρων ενέχει μεγαλύτερο βαθμό υποκειμενικότητας απ' ό,τι σε μια συστηματική ανασκόπηση με αυστηρό πρωτόκολλο. Αναγνώστες με εξειδικευμένο ενδιαφέρον σε ένα συγκεκριμένο υποπεδίο θα πρέπει να λάβουν αυτό υπόψη και να αναζητήσουν τις αντίστοιχες εξειδικευμένες ανασκοπήσεις της βιβλιογραφίας για βαθύτερη τεχνική ανάλυση.

1.8 Δεοντολογικές διαστάσεις της τεχνητής νοημοσύνης στις συστάσεις

Η ευρεία υιοθέτηση συστημάτων συστάσεων βασισμένων σε τεχνητή νοημοσύνη εγείρει ευρύτερα δεοντολογικά ερωτήματα που υπερβαίνουν τα καθαρά τεχνικά ζητήματα ακρίβειας ή απόδοσης. Από μια κριτική, κοινωνιολογική οπτική, η σχεδίαση αλγορίθμων σύστασης δεν αποτελεί ποτέ ουδέτερη τεχνική επιλογή, αλλά ενσωματώνει σιωπηρά αξιολογικές κρίσεις σχετικά με το ποιο περιεχόμενο αξίζει προβολή και ποιο όχι, με δυνητικές επιπτώσεις στη διαμόρφωση κοινής γνώμης και κοινωνικών αξιών.

Η παρούσα εργασία υιοθετεί μια τεχνικά εστιασμένη, αλλά δεοντολογικά ενημερωμένη οπτική: αναγνωρίζει ότι ζητήματα δικαιοσύνης, διαφάνειας και ιδιωτικότητας, που αναλύονται εκτενώς στο Κεφάλαιο 8, δεν αποτελούν δευτερεύοντα «πρόσθετα χαρακτηριστικά» ενός συστήματος σύστασης, αλλά θεμελιώδεις διαστάσεις σχεδίασης που πρέπει να ενσωματώνονται από το αρχικό στάδιο ανάπτυξης. Αυτή η θέση αντικατοπτρίζεται στη δομή της εργασίας, όπου τα ζητήματα δεοντολογίας και κοινωνικής ευθύνης διαπλέκονται με την τεχνική ανάλυση σε όλα τα κεφάλαια, αντί να συνοψίζονται αποκλειστικά σε ένα μεμονωμένο, «απομονωμένο» τμήμα.

1.9 Σχέση με συγγενικά ερευνητικά πεδία

Τα συστήματα συστάσεων διατηρούν στενή σχέση με γειτονικά ερευνητικά πεδία, χωρίς ωστόσο να ταυτίζονται πλήρως με κανένα από αυτά. Η ανάκτηση πληροφορίας (information retrieval) μοιράζεται με τα συστήματα συστάσεων τον στόχο της κατάταξης σχετικού περιεχομένου, αλλά διαφέρει στο ότι βασίζεται συνήθως σε ρητό ερώτημα του χρήστη (query), ενώ τα συστήματα συστάσεων λειτουργούν συχνά proactively, χωρίς ρητή έκφραση ανάγκης πληροφόρησης. Η μηχανική μάθηση και η βαθιά μάθηση, αντίθετα, αποτελούν εργαλειοθήκη τεχνικών που τα συστήματα συστάσεων αξιοποιούν, χωρίς να αποτελούν αυτά καθαυτά «πεδίο εφαρμογής», όπως αναλύθηκε εκτενώς στα Κεφάλαια 3 έως 5.

Η ανάλυση κοινωνικών δικτύων (social network analysis) παρέχει εργαλεία ανάλυσης γράφων που έχουν ενσωματωθεί στα GNN-based συστήματα σύστασης (Κεφάλαιο 5), ενώ η επεξεργασία φυσικής γλώσσας (NLP) τροφοδοτεί τεχνικές διήθησης βάσει περιεχομένου και τα σύγχρονα LLM-ενισχυμένα συστήματα εξηγήσεων (Ενότητα 8.5). Η σαφής χαρτογράφηση αυτών των διασυνδέσεων βοηθά στην κατανόηση του πεδίου των συστημάτων συστάσεων ως ενός διεπιστημονικού χώρου εφαρμογής, και όχι ως μιας απομονωμένης, αυτόνομης επιστημονικής κατηγορίας.

2. Θεωρητικό υπόβαθρο

Πριν από την ανάλυση των επιμέρους τεχνικών και αλγορίθμων που χρησιμοποιούνται στα σύγχρονα συστήματα συστάσεων, είναι απαραίτητο να τεθεί ένα σαφές θεωρητικό πλαίσιο. Στο πλαίσιο του παρόντος κεφαλαίου επιχειρείται η αποσαφήνιση των θεμελιωδών εννοιών του πεδίου, παρουσιάζει τις κατηγορίες των συστημάτων συστάσεων, περιγράφει τους τύπους δεδομένων που χρησιμοποιούνται και αναλύει τα κριτήρια αξιολόγησης της απόδοσής τους.

2.1 Ορισμός και βασικές έννοιες

Ένα σύστημα συστάσεων (recommendation system ή recommender system) ορίζεται ως ένα σύστημα φιλτραρίσματος πληροφοριών που στοχεύει στην πρόβλεψη της βαθμολογίας ή της προτίμησης που ένας χρήστης θα απέδιδε σε ένα αντικείμενο, και στη συνέχεια αξιοποιεί αυτή την πρόβλεψη ώστε να προτείνει αντικείμενα που ο χρήστης πιθανώς δεν έχει ακόμα αντιμετωπίσει αλλά θα τον ενδιέφεραν. Η θεμελιώδης παραδοχή είναι ότι οι αλληλεπιδράσεις χρηστών με αντικείμενα ακολουθούν νοηματικά μοτίβα τα οποία μπορούν να αναλυθούν και να αξιοποιηθούν για μελλοντικές προβλέψεις.

Τα τρία θεμελιώδη συστατικά στοιχεία ενός συστήματος συστάσεων είναι: (α) οι χρήστες (users), δηλαδή τα άτομα στα οποία απευθύνεται η σύσταση, (β) τα αντικείμενα (items), δηλαδή τα προϊόντα, ταινίες, τραγούδια, άρθρα ή οποιαδήποτε άλλη οντότητα προς σύσταση, και (γ) η συνάρτηση σύστασης (recommendation function), που αντιστοιχίζει ζεύγη χρήστη-αντικειμένου σε μια τιμή χρησιμότητας (utility value). Ο πίνακας αλληλεπίδρασης χρήστη-αντικειμένου (user-item interaction matrix ή utility matrix) αποτελεί την πρωταρχική δομή δεδομένων: κάθε στοιχείο r(u,i) εκφράζει τη βαθμολογία ή την αλληλεπίδραση του χρήστη u με το αντικείμενο i.

Ιδιαίτερη σημασία στη θεωρία των συστημάτων συστάσεων έχει η διάκριση μεταξύ ρητής (explicit) και έμμεσης (implicit) ανάδρασης. Η ρητή ανάδραση αφορά περιπτώσεις όπου ο χρήστης αξιολογεί άμεσα ένα αντικείμενο, για παράδειγμα με αστέρια ή αριθμητική βαθμολογία,

και αντιπροσωπεύει με μεγαλύτερη αξιοπιστία τις πραγματικές προτιμήσεις του χρήστη. Αντιθέτως, η έμμεση ανάδραση συλλέγεται έμμεσα από τη συμπεριφορά του χρήστη (κλικ, αγορές, χρόνος παραμονής σε μια σελίδα, ιστορικό αναζήτησης) και αποτελεί στην πράξη την κυρίαρχη μορφή δεδομένων, καθώς οι χρήστες σπάνια βαθμολογούν ρητά τα αντικείμενα με τα οποία αλληλεπιδρούν.

2.2 Ιστορική εξέλιξη των συστημάτων συστάσεων

Τα συστήματα συστάσεων δεν είναι φαινόμενο αποκλειστικά της σύγχρονης εποχής. Οι πρώτες ερευνητικές προσπάθειες εντοπίζονται ήδη από τα τέλη της δεκαετίας του 1980 και τις αρχές του 1990, με συστήματα όπως το Tapestry και το GroupLens που εφάρμοσαν πρωτόγονες μορφές collaborative filtering για φιλτράρισμα ηλεκτρονικών μηνυμάτων. Η ουσιαστική εφαρμογή τους στο εμπόριο ξεκίνησε στα μέσα της δεκαετίας του 1990 με την ανάπτυξη των ηλεκτρονικών αγορών Amazon και eBay, που αναγνώρισαν γρήγορα ότι μια ενιαία στρατηγική πρότασης προϊόντων δεν αρκούσε για να ικανοποιήσει τις διαφορετικές ανάγκες της πελατειακής τους βάσης.

Η δεκαετία 2000-2010 σηματοδοτήθηκε από τη σημαντική ανάπτυξη της Matrix Factorization (MF) ως τεχνικής αποσύνθεσης του πίνακα χρήστη-αντικειμένου σε λανθάνουσες αναπαραστάσεις χαμηλότερης διάστασης, ιδίως ύστερα από τον διαγωνισμό Netflix Prize (2006–2009), που ανέδειξε τις τεχνικές MF ως state-of-the-art για πρόβλεψη αξιολογήσεων. Η επόμενη μεγάλη στροφή ήρθε με τη διάδοση της βαθιάς μάθησης: από το 2016 και μετά, αρχιτεκτονικές όπως τα νευρωνικά δίκτυα συνέλιξης (Convolutional Neural Networks, CNNs), τα αναδρομικά νευρωνικά δίκτυα (Recurrent Neural Networks, RNNs) και τα αυτόματα κωδικοποιητικά (Autoencoders) άρχισαν να εφαρμόζονται με επιτυχία στο πεδίο.

Τα τελευταία χρόνια (2020 έως σήμερα) παρατηρείται μια ακόμα πιο δυναμική φάση εξέλιξης, με κυρίαρχες τάσεις την ενσωμάτωση αρχιτεκτονικών Transformer και μηχανισμών self-attention, τη χρήση Γραφικών Νευρωνικών Δικτύων (Graph Neural Networks, GNNs) που αναπαριστούν αλληλεπιδράσεις ως γράφους, και την ανάδυση τεχνικών αυτο-επιβλεπόμενης μάθησης (self-supervised learning) για την αντιμετώπιση της αραιότητας δεδομένων. Η εξέλιξη αυτή αντικατοπτρίζει τη συνεχή προσπάθεια της ερευνητικής κοινότητας να αντιμετωπίσει τα κλασικά προβλήματα του πεδίου με ολοένα πιο εκλεπτυσμένα εργαλεία.

2.3 Κατηγορίες συστημάτων συστάσεων

Η βιβλιογραφία αναγνωρίζει αρκετές διακριτές κατηγορίες συστημάτων συστάσεων, οι οποίες διαφέρουν ως προς τον τύπο των δεδομένων που αξιοποιούν και τον αλγόριθμο που εφαρμόζουν. Η ταξινόμηση αυτή δεν είναι αποκλειστική (πολλά σύγχρονα συστήματα συνδυάζουν στοιχεία από περισσότερες από μία κατηγορίες) αλλά είναι χρήσιμη ως εισαγωγικό πλαίσιο κατανόησης.

Η διήθηση βάσει περιεχομένου (content-based filtering) βασίζεται στα χαρακτηριστικά των αντικειμένων και στο ιστορικό προτιμήσεων του χρήστη. Η κεντρική ιδέα είναι απλή: αν ένας χρήστης άρεσε σε αντικείμενα με συγκεκριμένα χαρακτηριστικά στο παρελθόν, τότε πιθανόν να του αρέσουν και άλλα αντικείμενα με παρόμοια χαρακτηριστικά. Για την εξαγωγή χαρακτηριστικών από κείμενα χρησιμοποιείται συχνά η μέθοδος TF-IDF (Term Frequency-Inverse Document Frequency), ενώ η ομοιότητα υπολογίζεται συνήθως με κοσινουσιακή ομοιότητα.

Η συνεργατική διήθηση (collaborative filtering) αποτελεί ιστορικά την πιο διαδεδομένη οικογένεια αλγορίθμων. Στηρίζεται στη λογική ότι χρήστες με παρόμοιες προτιμήσεις στο παρελθόν θα έχουν παρόμοιες προτιμήσεις και στο μέλλον. Διακρίνεται σε user-based CF, που εντοπίζει χρήστες με παρόμοιο ιστορικό και προτείνει αντικείμενα που άρεσαν σε αυτούς, και item-based CF, που συγκρίνει αντικείμενα βάσει των βαθμολογιών που έχουν λάβει από διαφορετικούς χρήστες.

Τα υβριδικά συστήματα (hybrid systems) συνδυάζουν στοιχεία από δύο ή περισσότερες μεθόδους, με στόχο την εξουδετέρωση των αδυναμιών κάθε επιμέρους προσέγγισης. Τα συστήματα βασισμένα σε γνώση (knowledge-based systems) αξιοποιούν δομημένη γνώση για τη δημιουργία συστάσεων, ιδιαίτερα αποτελεσματικά σε πεδία με πολύπλοκες απαιτήσεις, όπως η υγεία ή τα νομικά συστήματα. Τέλος, τα συστήματα ευαίσθητα στο πλαίσιο (context-aware systems) ενσωματώνουν πλαισιακές πληροφορίες όπως τοποθεσία, ώρα, καιρός ή συναισθηματική κατάσταση, ώστε να παρέχουν δυναμικές και εξατομικευμένες συστάσεις.

Πίνακας 2.1: Συνοπτική κατηγοριοποίηση συστημάτων συστάσεων

Κατηγορία	Αρχή λειτουργίας	Κύριο πλεονέκτημα
Content-based	Σύγκριση χαρακτηριστικών αντικειμένων	Χωρίς ανάγκη άλλων χρηστών
Collaborative Filtering	Εύρεση ομοιότητας μεταξύ χρηστών/αντικειμένων	Αξιοποίηση κοινωνικής πληροφορίας
Hybrid	Συνδυασμός παραπάνω μεθόδων	Ανθεκτικότητα σε αδυναμίες
Knowledge-based	Χρήση γνωσιακού γράφου	Δουλεύει με λίγα δεδομένα
Context-aware	Ενσωμάτωση πλαισιακής πληροφορίας	Δυναμική εξατομίκευση
Deep Learning-based	Νευρωνικά δίκτυα για αναπαράσταση	Σύλληψη μη γραμμικών σχέσεων

2.4 Δεδομένα και αναπαράσταση χρηστών-αντικειμένων

Η ποιότητα ενός συστήματος συστάσεων εξαρτάται άμεσα από την ποιότητα και την ποσότητα των διαθέσιμων δεδομένων. Στην πράξη, ο πίνακας χρήστη-αντικειμένου είναι τυπικά εξαιρετικά αραιός (sparse): ακόμα και σε μεγάλες πλατφόρμες, κάθε χρήστης έχει αλληλεπιδράσει μόνο με ένα μικρό κλάσμα των διαθέσιμων αντικειμένων. Αυτή η αραιότητα αποτελεί έναν από τους κεντρικούς προκλήσεις του πεδίου, καθώς περιορίζει σοβαρά την ικανότητα πρόβλεψης των αλγορίθμων.

Η αναπαράσταση χρηστών και αντικειμένων ως διανύσματα χαμηλής διάστασης (embeddings) αποτελεί θεμελιώδη τεχνική στα σύγχρονα συστήματα συστάσεων. Ο χρήστης u αναπαρίσταται ως ένα διάνυσμα $p_u \in R^k$ και κάθε αντικείμενο i ως $q_i \in R^k$, όπου k είναι ο αριθμός των λανθανουσών διαστάσεων (latent dimensions). Η προβλεπόμενη βαθμολογία υπολογίζεται τότε ως το εσωτερικό γινόμενο των δύο διανυσμάτων: $\hat{r}(u,i) = p_u \cdot q_i$. Αυτή η προσέγγιση, γνωστή ως Matrix Factorization, αποτελεί τη βάση πάνω στην οποία οικοδομήθηκαν στη συνέχεια οι αρχιτεκτονικές βαθιάς μάθησης.

Σε πλαίσια γραφικής αναπαράστασης, οι χρήστες και τα αντικείμενα μοντελοποιούνται ως κόμβοι ενός γράφου, ενώ οι αλληλεπιδράσεις αναπαρίστανται ως ακμές. Αυτή η προσέγγιση επιτρέπει την εκμετάλλευση της δομικής πληροφορίας του γράφου (κοινές συνδέσεις, γειτονικότητα υψηλής τάξης) για τη βελτίωση των αναπαραστάσεων, κάτι που αποτελεί τη βάση λειτουργίας των Γραφικών Νευρωνικών Δικτύων.

2.5 Μετρικές αξιολόγησης

Η αξιολόγηση της απόδοσης ενός συστήματος συστάσεων αποτελεί κρίσιμο στάδιο τόσο κατά την έρευνα όσο και κατά την ανάπτυξη εφαρμογών. Οι μετρικές που χρησιμοποιούνται διαιρούνται γενικά σε δύο κατηγορίες: τις μετρικές ακρίβειας πρόβλεψης (prediction accuracy metrics), που μετράνε πόσο κοντά βρίσκεται η πρόβλεψη του συστήματος στη βαθμολογία που έδωσε πραγματικά ο χρήστης, και τις μετρικές ταξινόμησης (ranking metrics), που αξιολογούν την ποιότητα της λίστας συστάσεων που παράγει το σύστημα.

Η μέση απόλυτη απόκλιση (Mean Absolute Error, MAE) και η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (Root Mean Squared Error, RMSE) αποτελούν τις πλέον διαδεδομένες μετρικές ακρίβειας πρόβλεψης. Η RMSE τιμωρεί περισσότερο τα μεγάλα σφάλματα σε σχέση με τη MAE, γεγονός που την καθιστά προτιμότερη όταν οι ακραίες αποκλίσεις έχουν ιδιαίτερη σημασία. Για σενάρια top-N σύστασης, όπου το ζητούμενο είναι η παραγωγή μιας λίστας με τα N πιο σχετικά αντικείμενα, χρησιμοποιούνται μετρικές όπως η Precision@K, η Recall@K, η Normalized Discounted Cumulative Gain (NDCG@K) και η Hit Rate@K. Η NDCG αποτελεί

ιδιαίτερα εκφραστική μετρική, καθώς λαμβάνει υπόψη και τη θέση κάθε σχετικού αντικειμένου στη λίστα, αποδίδοντας μεγαλύτερο βάρος στα αντικείμενα που εμφανίζονται ψηλότερα.

Πίνακας 2.2: Μετρικές αξιολόγησης συστημάτων συστάσεων

Μετρική	Ορισμός	Χρήση
MAE	Μέσο απόλυτο σφάλμα πρόβλεψης βαθμολογίας	Αξιολόγηση ακρίβειας πρόβλεψης
RMSE	Τετραγωνική ρίζα μέσου τετραγωνικού σφάλματος	Τιμωρεί μεγάλα σφάλματα
Precision@K	Αναλογία σχετικών αντικειμένων στα K πρώτα	Ranking-based αξιολόγηση
Recall@K	Κάλυψη σχετικών αντικειμένων στα K πρώτα	Ranking-based αξιολόγηση
NDCG@K	Κανονικοποιημένο κέρδος βαθμολόγησης κατάταξης	Ποιότητα ταξινόμησης λίστας
Hit Rate@K	Ποσοστό χρηστών με σχετικό αντικείμενο στα K πρώτα	Top-N σύσταση
AUC	Εμβαδόν κάτω από καμπύλη ROC	Δυαδική κατάταξη

2.6 Κεντρικά προβλήματα του πεδίου

Ανεξαρτήτως της μεθόδου που χρησιμοποιείται, τα συστήματα συστάσεων αντιμετωπίζουν ορισμένα θεμελιώδη προβλήματα που η ερευνητική κοινότητα προσπαθεί να αντιμετωπίσει εδώ και δεκαετίες. Η κατανόηση αυτών των προβλημάτων είναι απαραίτητη για την εκτίμηση των επιλογών σχεδιασμού που κάνουν οι ερευνητές.

Το πρόβλημα της αραιότητας δεδομένων (data sparsity) αναφέρεται στο γεγονός ότι οι περισσότεροι χρήστες έχουν αλληλεπιδράσει μόνο με ένα ελάχιστο ποσοστό των διαθέσιμων αντικειμένων, αφήνοντας τον πίνακα χρήστη-αντικειμένου κατά κύριο λόγο κενό. Αυτό δυσχεραίνει σοβαρά την εκπαίδευση αλγορίθμων και οδηγεί σε αναξιόπιστες προβλέψεις για αντικείμενα με λίγες αξιολογήσεις.

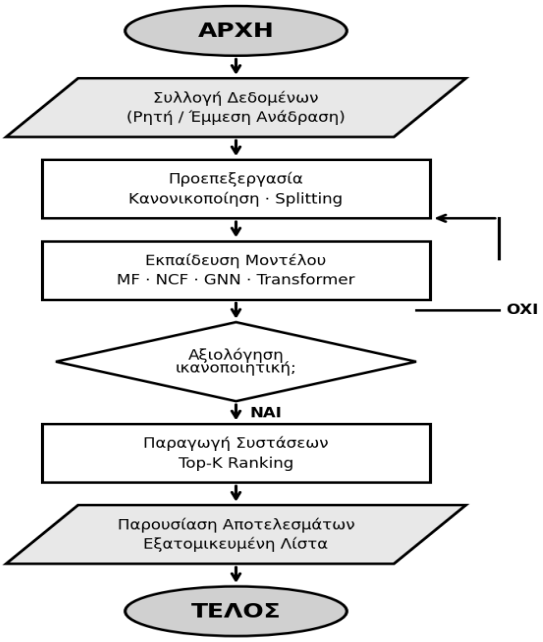
Το πρόβλημα του νέου χρήστη και του νέου αντικειμένου (cold start) εμφανίζεται όταν ένας νέος χρήστης ή ένα νέο αντικείμενο εισέρχεται στο σύστημα χωρίς ιστορικό αλληλεπίδρασης. Σε αυτή την περίπτωση, το σύστημα δεν διαθέτει επαρκείς πληροφορίες για να παράγει αξιόπιστες συστάσεις. Το πρόβλημα αυτό αποτελεί ένα από τα πιο επίμονα ζητήματα του πεδίου και έχει ωθήσει στην ανάπτυξη τεχνικών όπως η χρήση γνωσιακών γράφων, τα embeddings μεταφοράς γνώσης (transfer learning embeddings) και τα συστήματα ευαίσθητα στο πλαίσιο.

Η προκατάληψη δημοτικότητας (popularity bias) αναφέρεται στην τάση των συστημάτων να ευνοούν δημοφιλή αντικείμενα εις βάρος λιγότερο γνωστών, αλλά εξίσου σχετικών για τον συγκεκριμένο χρήστη. Αυτό οδηγεί σε μειωμένη ποικιλομορφία (diversity) και νέο αποκαλούμενη «φυσαλίδα φίλτρων» (filter bubble), όπου ο χρήστης εκτίθεται μόνο σε ένα στενό εύρος αντικειμένων. Τέλος, η κλιμακωσιμότητα (scalability) αποτελεί πρόκληση για εφαρμογές πραγματικού κόσμου με εκατομμύρια χρήστες και αντικείμενα, καθώς πολλοί αλγόριθμοι αδυνατούν να ανταποκριθούν αποτελεσματικά σε αυτή την κλίμακα.

2.7 Διαδικασία λειτουργίας ενός συστήματος συστάσεων

Το Σχήμα 2.1 απεικονίζει τη γενική διαδικασία λειτουργίας ενός συστήματος συστάσεων, από τη συλλογή δεδομένων αλληλεπίδρασης έως την παρουσίαση της εξατομικευμένης λίστας στον χρήστη. Κεντρικό στοιχείο είναι ο βρόχος αξιολόγησης: αν το μοντέλο δεν ικανοποιεί τα κριτήρια απόδοσης, επιστρέφει στο στάδιο εκπαίδευσης για βελτιστοποίηση. Η διαδικασία επαναλαμβάνεται έως ότου επιτευχθεί ικανοποιητική απόδοση σε validation set.

Σχήμα 2.1: Γενική Διαδικασία Συστήματος Συστάσεων



Σχήμα 2.1: Γενική διαδικασία λειτουργίας συστήματος συστάσεων

2.8 Αριθμητικό παράδειγμα συνεργατικής διήθησης

Για να γίνει κατανοητή η λειτουργία της συνεργατικής διήθησης, παρουσιάζεται ένα πλήρες αριθμητικό παράδειγμα με πέντε χρήστες και πέντε ταινίες. Ο Πίνακας 2.3 δείχνει τον πίνακα αλληλεπίδρασης χρήστη-αντικειμένου, όπου κάθε στοιχείο αναπαριστά τη βαθμολογία (1-5) που έδωσε ο χρήστης στην ταινία, και το σύμβολο «: » υποδηλώνει ότι ο χρήστης δεν έχει αξιολογήσει την ταινία. Στόχος είναι να προβλεφθούν οι βαθμολογίες της νέας χρήστριας Ελένης.

Πίνακας 2.3: Παράδειγμα πίνακα αλληλεπίδρασης χρήστη-ταινίας

Χρήστης	Matrix (1-5)	Inception	Titanic	Avatar	Forrest G.
Άλκης	5	4	—	3	—
Βίκυ	4	—	5	4	3
Γιώργος	—	3	4	—	5
Δήμητρα	5	4	—	4	—
Ελένη (νέα)	4	?	?	?	?

Για την εύρεση των πλησιέστερων γειτόνων της Ελένης, υπολογίζεται η κοσινουσιακή ομοιότητα μεταξύ του διανύσματός της και των υπολοίπων χρηστών. Η Ελένη έχει βαθμολογήσει μόνο τη Matrix (4), οπότε το διάνυσμά της είναι [4, 0, 0, 0, 0]. Ο Άλκης έχει βαθμολογήσει επίσης τη Matrix (5): sim(Ελένη, Άλκης) = (4×5)/(√16 × √(25+16+9)) = 20/(4×√50) = 20/28.28 ≈ 0.71. Η Δήμητρα έχει επίσης Matrix=5: sim(Ελένη, Δήμητρα) = (4×5)/(4×√(25+16+16)) ≈ 0.67. Βάσει αυτών των ομοιοτήτων, οι πλησιέστεροι γείτονες είναι Άλκης και Δήμητρα, οπότε οι συστάσεις για την Ελένη θα περιλαμβάνουν Inception (4+4)/2=4 και Avatar (3+4)/2=3.5: ταινίες που άρεσαν σε χρήστες με παρόμοιες προτιμήσεις.

2.9 Τεχνικές χειρισμού της implicit feedback

Στην πράξη, τα περισσότερα πραγματικά συστήματα συστάσεων βασίζονται σε έμμεση ανάδραση (implicit feedback) (κλικ, προβολές, χρόνο παραμονής, αγορές) και όχι σε ρητές βαθμολογίες, καθώς οι χρήστες σπάνια αφιερώνουν χρόνο για να βαθμολογήσουν ρητά αντικείμενα. Η έμμεση ανάδραση, ωστόσο, παρουσιάζει δύο θεμελιώδεις διαφορές σε σχέση με τη ρητή: πρώτον, είναι μονόπλευρα θετική, καθώς η απουσία αλληλεπίδρασης δεν σημαίνει απαραίτητα αρνητική προτίμηση, αλλά μπορεί απλώς να αντικατοπτρίζει άγνοια ή έλλειψη έκθεσης στο αντικείμενο (exposure bias). Δεύτερον, εμπεριέχει διαφορετικά επίπεδα εμπιστοσύνης ανά παρατήρηση, καθώς μια αγορά αποτελεί ισχυρότερη ένδειξη προτίμησης από ένα απλό κλικ.

Για τον χειρισμό αυτών των ιδιαιτεροτήτων έχουν αναπτυχθεί ειδικές τεχνικές. Η αρνητική δειγματοληψία (negative sampling) επιλέγει τυχαία ή με ευρετικά κριτήρια αντικείμενα με τα οποία ο χρήστης δεν έχει αλληλεπιδράσει, τα οποία θεωρούνται προσωρινά ως «αρνητικά» παραδείγματα κατά την εκπαίδευση, όπως ακριβώς συμβαίνει στη συνάρτηση κόστους BPR. Η σταθμισμένη ALS για έμμεση ανάδραση (Weighted ALS) εισάγει έναν συντελεστή εμπιστοσύνης $c_{ui} = 1 + \alpha \cdot r_{ui}$ στη συνάρτηση κόστους, ώστε οι παρατηρήσεις με υψηλότερη συχνότητα αλληλεπίδρασης να βαρύνουν περισσότερο στη βελτιστοποίηση. Τέλος, οι μέθοδοι απανθρωπισμού της έκθεσης (exposure debiasing) προσπαθούν να διαχωρίσουν τη πραγματική προτίμηση του χρήστη από την επίδραση του ίδιου του συστήματος σύστασης στο ποια αντικείμενα είδε ο χρήστης εξαρχής, ένα πρόβλημα γνωστό ως feedback loop bias.

2.10 Μετρικές ομοιότητας: αναλυτική σύγκριση

Η επιλογή της μετρικής ομοιότητας αποτελεί κρίσιμη σχεδιαστική απόφαση στις μεθόδους που βασίζονται σε γειτονιά (neighborhood-based), όπως η συνεργατική διήθηση που παρουσιάστηκε στο Κεφάλαιο 3. Η ομοιότητα συνημιτόνου (cosine similarity) μετρά τη γωνία μεταξύ δύο διανυσμάτων αξιολόγησης, αγνοώντας τη συνολική κλίμακα βαθμολόγησης κάθε χρήστη, γεγονός που τη κάνει ευάλωτη σε συστηματικές διαφορές στο πώς οι χρήστες χρησιμοποιούν την κλίμακα βαθμολόγησης (π.χ. ένας χρήστης που βαθμολογεί πάντα αυστηρά).

Ο συντελεστής συσχέτισης Pearson αντιμετωπίζει αυτό το πρόβλημα αφαιρώντας τον μέσο όρο βαθμολόγησης κάθε χρήστη πριν τον υπολογισμό της ομοιότητας, καθιστώντας τη μετρική ανεξάρτητη από συστηματικές μεροληψίες κλίμακας. Ο δείκτης Jaccard, αντίθετα, αγνοεί πλήρως τις τιμές βαθμολόγησης και εστιάζει αποκλειστικά στο ποια αντικείμενα έχουν αξιολογηθεί από κοινού, καθιστώντας τον ιδιαίτερα χρήσιμο σε σενάρια implicit feedback (Ενότητα 2.9), όπου δεν υπάρχουν καν αριθμητικές βαθμολογίες παρά μόνο δυαδικές ενδείξεις αλληλεπίδρασης.

Η προσαρμοσμένη ομοιότητα συνημιτόνου (adjusted cosine similarity) συνδυάζει τα πλεονεκτήματα των δύο πρώτων προσεγγίσεων, εφαρμόζοντας κεντροποίηση ως προς τον μέσο όρο του χρήστη πριν τον υπολογισμό της ομοιότητας συνημιτόνου μεταξύ αντικειμένων, και έχει αποδειχθεί εμπειρικά ανώτερη της απλής ομοιότητας συνημιτόνου στο πλαίσιο της item-based συνεργατικής διήθησης (Ενότητα 3.2).

2.11 Ταξινόμηση του προβλήματος ψυχρής εκκίνησης (cold start)

Το πρόβλημα της cold start, που αναφέρθηκε συνοπτικά στην Ενότητα 2.6, διακρίνεται σε τρεις βασικές κατηγορίες ανάλογα με την πηγή της έλλειψης δεδομένων. Η ψυχρή εκκίνηση νέου χρήστη (new user cold start) αφορά χρήστες χωρίς ιστορικό αλληλεπίδρασης, για τους οποίους το σύστημα δεν διαθέτει κανένα σήμα προτίμησης. Η ψυχρή εκκίνηση νέου αντικειμένου (new item cold start) αφορά αντικείμενα που μόλις προστέθηκαν στην πλατφόρμα και δεν έχουν λάβει καμία αλληλεπίδραση, καθιστώντας αδύνατο τον υπολογισμό λανθάνοντων παραγόντων μέσω παραγοντοποίησης πίνακα.

Η τρίτη κατηγορία, η ψυχρή εκκίνηση συστήματος (system cold start), εμφανίζεται όταν ολόκληρη η πλατφόρμα είναι νέα και δεν διαθέτει επαρκή ιστορικά δεδομένα αλληλεπίδρασης για κανέναν χρήστη ή αντικείμενο, καθιστώντας αναποτελεσματικές τις περισσότερες μεθόδους συνεργατικής διήθησης. Σε αυτή την περίπτωση, οι υβριδικές προσεγγίσεις που συνδυάζουν διήθηση βάσει περιεχομένου με δημογραφικά χαρακτηριστικά αποτελούν συχνά τη μόνη βιώσιμη λύση, έως ότου συσσωρευτεί επαρκής όγκος δεδομένων αλληλεπίδρασης ώστε να ενεργοποιηθούν αποτελεσματικά οι τεχνικές συνεργατικής διήθησης.

2.12 Διαστάσεις αξιολόγησης πέραν της ακρίβειας

Πέραν των μετρικών ακρίβειας πρόβλεψης που αναφέρθηκαν στην Ενότητα 2.5 (Precision@K, Recall@K, NDCG@K), η σύγχρονη βιβλιογραφία αναγνωρίζει ότι ένα ολοκληρωμένο σύστημα σύστασης πρέπει να αξιολογείται και ως προς πρόσθετες, εξίσου σημαντικές διαστάσεις. Η ποικιλομορφία (diversity) μετρά τον βαθμό στον οποίο η προτεινόμενη λίστα περιλαμβάνει αντικείμενα από διαφορετικές κατηγορίες ή με διαφορετικά χαρακτηριστικά, αντί να επικεντρώνεται σε ένα στενό υποσύνολο πολύ όμοιων αντικειμένων.

Η νεωτερικότητα (novelty) και η σπανιότητα (serendipity) αποτιμούν κατά πόσο οι συστάσεις εκθέτουν τον χρήστη σε αντικείμενα που δεν θα είχε ανακαλύψει μόνος του, πέραν των ήδη δημοφιλών ή προφανών επιλογών: διάσταση που συνδέεται άμεσα με το πρόβλημα της προκατάληψης δημοτικότητας (popularity bias) που αναφέρθηκε στην Ενότητα 2.9. Η κάλυψη καταλόγου (catalog coverage), τέλος, μετρά το ποσοστό του συνολικού καταλόγου αντικειμένων που εμφανίζεται ποτέ σε προτάσεις προς οποιονδήποτε χρήστη, αποτυπώνοντας κατά πόσο το σύστημα «εγκλωβίζεται» σε ένα στενό υποσύνολο δημοφιλών αντικειμένων εις βάρος της long-tail κατανομής περιεχομένου.

Η ταυτόχρονη βελτιστοποίηση ακρίβειας και αυτών των συμπληρωματικών διαστάσεων δημιουργεί συχνά αντικρουόμενους στόχους: η μεγιστοποίηση της ακρίβειας τείνει να ευνοεί δημοφιλή, «ασφαλή» αντικείμενα, ενώ η μεγιστοποίηση της ποικιλομορφίας ή της σπανιότητας μπορεί να υποβαθμίσει τη μετρούμενη ακρίβεια. Τεχνικές επανακατάταξης (re-ranking) με ρητούς περιορισμούς ποικιλομορφίας, παρόμοιες με αυτές που αναφέρθηκαν στον Πίνακα 8.2 για τη δικαιοσύνη, αποτελούν την κυριότερη πρακτική προσέγγιση εξισορρόπησης αυτών των αντικρουόμενων στόχων.

2.13 Αναπαράσταση δεδομένων χρόνου και εποχικότητας

Πέραν της βασικής αναπαράστασης αλληλεπιδράσεων χρήστη-αντικειμένου που παρουσιάστηκε στην Ενότητα 2.4, πολλές πραγματικές εφαρμογές απαιτούν ρητή μοντελοποίηση της χρονικής διάστασης. Η χρονοσφραγίδα (timestamp) μιας αλληλεπίδρασης μπορεί να αξιοποιηθεί με διάφορους τρόπους: ως απλό φίλτρο πρόσφατων δεδομένων (recency weighting), όπου παλαιότερες αλληλεπιδράσεις λαμβάνουν προοδευτικά μικρότερο βάρος στη συνάρτηση κόστους, ή ως ρητό χαρακτηριστικό εισόδου σε αρχιτεκτονικές σειριακής σύστασης (Κεφάλαιο 6.1), όπου η σχετική διαφορά χρόνου μεταξύ διαδοχικών αλληλεπιδράσεων κωδικοποιείται απευθείας.

Η εποχικότητα (seasonality) αποτελεί ειδική περίπτωση χρονικής μεταβλητότητας, ιδιαίτερα σημαντική σε τομείς όπως ο τουρισμός, το λιανικό εμπόριο ή η γεωργία (Ενότητα 7.10), όπου οι προτιμήσεις των χρηστών παρουσιάζουν επαναλαμβανόμενα μοτίβα συνδεδεμένα με την εποχή του έτους ή ειδικές περιόδους (π.χ. εορτές). Μοντέλα context-aware

σύστασης (Ενότητα 6.3) ενσωματώνουν συνήθως την εποχικότητα ως ρητή πλαισιακή μεταβλητή, ενώ πιο σύγχρονες προσεγγίσεις αξιοποιούν περιοδικές συναρτήσεις κωδικοποίησης θέσης, παρόμοιες με αυτές του μηχανισμού Transformer (Ενότητα 4.12), για την αποτύπωση κυκλικών χρονικών μοτίβων.

2.14 Πρωτόκολλα διαχωρισμού δεδομένων εκπαίδευσης/δοκιμής

Η αξιόπιστη αξιολόγηση ενός μοντέλου σύστασης απαιτεί προσεκτικό διαχωρισμό των διαθέσιμων δεδομένων σε σύνολα εκπαίδευσης και δοκιμής, με τρόπο που να αποφεύγει τη διαρροή πληροφορίας (data leakage) από το ένα σύνολο στο άλλο. Ο τυχαίος διαχωρισμός (random split), όπου κάθε αλληλεπίδραση ανεξάρτητα ανατίθεται τυχαία σε ένα από τα δύο σύνολα, είναι η απλούστερη προσέγγιση, αλλά αγνοεί τη χρονική διάσταση των δεδομένων, επιτρέποντας ενδεχομένως στο μοντέλο να «δει» μελλοντικές αλληλεπιδράσεις κατά την εκπαίδευση και να αξιολογηθεί σε παλαιότερες: σενάριο μη ρεαλιστικό για πρακτική εφαρμογή.

Ο χρονικός διαχωρισμός (temporal/leave-last-out split) αντιμετωπίζει αυτό το πρόβλημα διατηρώντας τις χρονικά τελευταίες αλληλεπιδράσεις κάθε χρήστη (ή ολόκληρου του dataset) ως σύνολο δοκιμής, εξασφαλίζοντας ότι η αξιολόγηση αντικατοπτρίζει το ρεαλιστικό σενάριο πρόβλεψης μελλοντικής συμπεριφοράς βάσει παρελθοντικού ιστορικού: ιδιαίτερα σημαντικό για τη δίκαιη αξιολόγηση σειριακών μοντέλων (Κεφάλαιο 6.1). Η επιλογή πρωτοκόλλου διαχωρισμού μπορεί να επηρεάσει σημαντικά τη σχετική κατάταξη μοντέλων σε συγκριτικές αξιολογήσεις, αποτελώντας πρόσθετη πηγή μεθοδολογικής μεταβλητότητας πέραν αυτής που συζητήθηκε στην Ενότητα 9.9.

2.15 Σύνοψη

Αυτό το κεφάλαιο έθεσε τα θεμέλια για την κατανόηση του πεδίου των συστημάτων συστάσεων. Ορίστηκαν οι βασικές έννοιες (χρήστες, αντικείμενα, πίνακας αλληλεπίδρασης, ρητή και έμμεση ανάδραση) και παρουσιάστηκε η ιστορική εξέλιξη του πεδίου από τις κλασικές μεθόδους έως τις σύγχρονες αρχιτεκτονικές βαθιάς μάθησης. Η ταξινόμηση των κατηγοριών συστημάτων συστάσεων, η ανάλυση των τύπων δεδομένων, οι μετρικές αξιολόγησης και τα κεντρικά προβλήματα του πεδίου αποτελούν το αναγκαίο υπόβαθρο για την κατανόηση των αναλύσεων που ακολουθούν στα επόμενα κεφάλαια.

3. Κλασικές μέθοδοι σύστασης

Πριν από την έλευση της βαθιάς μάθησης, τα συστήματα συστάσεων βασίζονταν σε μια οικογένεια αλγορίθμων που σήμερα αναφέρονται συλλογικά ως «κλασικές» ή «παραδοσιακές» μέθοδοι. Οι μέθοδοι αυτές δεν έχουν μόνο ιστορική αξία: αποτελούν ακόμα σήμερα τη βάση πάνω στην οποία οικοδομούνται πολλά υβριδικά και νευρωνικά συστήματα, και η κατανόησή τους είναι απαραίτητη για την αξιολόγηση της εξέλιξης του πεδίου. Στις επόμενες ενότητες εξετάζονται η διήθηση βάσει περιεχομένου, η συνεργατική διήθηση και οι παραλλαγές της, οι τεχνικές παραγοντοποίησης πίνακα και τα υβριδικά συστήματα.

3.1 Διήθηση βάσει περιεχομένου (Content-based filtering)

Η διήθηση βάσει περιεχομένου (content-based filtering, CBF) αποτελεί ιστορικά την πρώτη από τις μεγάλες κατηγορίες αλγορίθμων σύστασης. Η βασική ιδέα είναι απλή και διαισθητική: αν ένας χρήστης έχει δείξει ενδιαφέρον για αντικείμενα με ορισμένα χαρακτηριστικά στο παρελθόν, τότε πιθανότατα θα τον ενδιαφέρουν και άλλα αντικείμενα με παρόμοια χαρακτηριστικά. Η μέθοδος λειτουργεί αποκλειστικά βάσει της περιγραφής των αντικειμένων και του ιστορικού του συγκεκριμένου χρήστη, χωρίς να χρειάζεται πληροφορία από άλλους χρήστες.

Η υλοποίηση μιας τέτοιας προσέγγισης περιλαμβάνει τρία βασικά στάδια. Πρώτον, την εξαγωγή χαρακτηριστικών (feature extraction) από τα αντικείμενα: για έγγραφα και κείμενα χρησιμοποιείται κατά κανόνα η μέθοδος TF-IDF (Term Frequency-Inverse Document Frequency), η οποία αποδίδει βάρος σε κάθε λέξη ανάλογα με τη συχνότητά της στο έγγραφο και αντίστροφα ανάλογα με τη συχνότητά της στο σύνολο των εγγράφων. Δεύτερον, τη δόμηση του προφίλ χρήστη (user profile), που εκφράζει τις προτιμήσεις του χρήστη ως σταθμισμένο μέσο των χαρακτηριστικών διανυσμάτων των αντικειμένων με τα οποία έχει αλληλεπιδράσει θετικά. Τρίτον, τον υπολογισμό ομοιότητας μεταξύ του προφίλ χρήστη και κάθε υποψήφιου αντικειμένου, συνήθως με κοσινουσιακή ομοιότητα.

3.1.1 Πλεονεκτήματα και μειονεκτήματα

Το κύριο πλεονέκτημα της CBF είναι ότι δεν εξαρτάται από δεδομένα άλλων χρηστών, κάτι που την καθιστά ανθεκτική στα προβλήματα αραιότητας που ταλαιπωρούν τη συνεργατική διήθηση. Επίσης, οι συστάσεις της είναι ερμηνεύσιμες: το σύστημα μπορεί να εξηγήσει γιατί πρότεινε ένα συγκεκριμένο αντικείμενο, αναφερόμενο στα κοινά χαρακτηριστικά με αντικείμενα που άρεσαν στον χρήστη στο παρελθόν. Αυτό έχει σημαντική αξία σε εφαρμογές που απαιτούν διαφάνεια.

Ωστόσο, η CBF παρουσιάζει και σημαντικές αδυναμίες. Το φαινόμενο της υπερεξειδίκευσης (over-specialization) αναφέρεται στην τάση του συστήματος να προτείνει αντικείμενα πολύ παρόμοια με αυτά που ο χρήστης έχει ήδη δει, μειώνοντας έτσι την ποικιλομορφία και το στοιχείο της έκπληξης στις συστάσεις. Επίσης, για έναν νέο χρήστη χωρίς ιστορικό αλληλεπίδρασης, το σύστημα δεν μπορεί να παράγει αξιόπιστες συστάσεις: πρόβλημα γνωστό ως cold start νέου χρήστη. Τέλος, η CBF δυσκολεύεται να χειριστεί πλούσια πολυμεσικά αντικείμενα, όπως εικόνες ή βίντεο, χωρίς τη βοήθεια εξελιγμένων τεχνικών εξαγωγής χαρακτηριστικών.

3.2 Συνεργατική διήθηση (Collaborative filtering)

Η συνεργατική διήθηση (collaborative filtering, CF) αποτελεί ιστορικά την πιο διαδεδομένη και μελετημένη κατηγορία αλγορίθμων σύστασης. Η ιδέα της εισήχθη το 1992 από τους Goldberg, Nichols, Oki και Terry στο σύστημα Tapestry, αρχικά για φιλτράρισμα ηλεκτρονικής αλληλογραφίας, και αναπτύχθηκε ραγδαία τα επόμενα χρόνια. Σε αντίθεση με την CBF, η CF δεν χρειάζεται πληροφορία για τα χαρακτηριστικά των αντικειμένων: βασίζεται αποκλειστικά στις αλληλεπιδράσεις χρηστών-αντικειμένων, αξιοποιώντας τη λογική ότι χρήστες με παρόμοιο ιστορικό στο παρελθόν θα έχουν παρόμοιες προτιμήσεις και στο μέλλον.

Η CF διακρίνεται σε δύο μεγάλες υποκατηγορίες ανάλογα με τη μονάδα σύγκρισης που χρησιμοποιεί: τη μνημονική CF (memory-based CF) και τη μοντελο-βασισμένη CF (model-based CF). Η μνημονική CF εργάζεται άμεσα με τον πίνακα αλληλεπίδρασης χρηστή-αντικειμένου, υπολογίζοντας ομοιότητες είτε μεταξύ χρηστών (user-based) είτε μεταξύ αντικειμένων (item-based). Η μοντελο-βασισμένη CF εκπαιδεύει ένα στατιστικό ή μηχανιστικό μοντέλο στα δεδομένα αλληλεπίδρασης, χρησιμοποιώντας στη συνέχεια το μοντέλο για πρόβλεψη.

3.2.1 User-based συνεργατική διήθηση

Στη user-based CF, ο αλγόριθμος αναζητά χρήστες που έχουν παρόμοιο ιστορικό αλληλεπίδρασης με τον στοχευόμενο χρήστη u. Η ομοιότητα μεταξύ δύο χρηστών υπολογίζεται συνήθως με κοσινουσιακή ομοιότητα ή με τον συντελεστή συσχέτισης Pearson. Μόλις εντοπιστεί το σύνολο των k πλησιέστερων γειτόνων (k-nearest neighbors, kNN), η προβλεπόμενη βαθμολογία του χρήστη u για ένα αντικείμενο i υπολογίζεται ως σταθμισμένος μέσος όρος των βαθμολογιών που έδωσαν οι γείτονες στο ίδιο αντικείμενο, με βάρη ανάλογα της ομοιότητάς τους με τον u.

Η user-based CF παρουσιάζει πρακτικά μειονεκτήματα σε εφαρμογές μεγάλης κλίμακας. Ο υπολογισμός ομοιότητας μεταξύ όλων των ζευγών χρηστών έχει πολυπλοκότητα $O(|U|^2)$, κάτι που γίνεται απαγορευτικό όταν ο αριθμός χρηστών $|U|$ είναι της τάξης των εκατομμυρίων. Επιπλέον, τα προφίλ χρηστών αλλάζουν συνεχώς, απαιτώντας συχνή επανυπολογισμό ομοιοτήτων. Αυτά τα μειονεκτήματα ώθησαν στην ανάπτυξη της item-based CF ως πιο αποτελεσματικής εναλλακτικής για πολλές πρακτικές εφαρμογές.

3.2.2 Item-based συνεργατική διήθηση

Η item-based CF, που αναπτύχθηκε και εφαρμόστηκε με επιτυχία στη μεγάλη κλίμακα από την Amazon, μετατοπίζει τη μονάδα σύγκρισης από τους χρήστες στα αντικείμενα. Η βασική παρατήρηση είναι ότι τα αντικείμενα τείνουν να αξιολογούνται παρόμοια από χρήστες που τα αξιολόγησαν και τα δύο, οπότε η ομοιότητα αντικειμένων είναι πιο σταθερή στο χρόνο σε σχέση με τη χρήστη-σε-χρήστη ομοιότητα. Για να γίνει μια σύσταση στον χρήστη u, το σύστημα εντοπίζει αντικείμενα παρόμοια με αυτά που ο u έχει αξιολογήσει θετικά και τα προτείνει ως επόμενες επιλογές.

Η item-based CF προσφέρει σημαντικό πλεονέκτημα κλιμακωσιμότητας (scalability): οι ομοιότητες αντικειμένων μπορούν να προϋπολογιστούν εκτός σύνδεσης (offline) και να

αποθηκευτούν, με αποτέλεσμα ο χρόνος σύστασης σε πραγματικό χρόνο να είναι πολύ μικρός. Αυτό την κατέστησε την επιλογή της Amazon για τη λειτουργία του «Customers who bought this also bought», μιας από τις πιο επιτυχημένες εφαρμογές συστημάτων συστάσεων στην ιστορία του ηλεκτρονικού εμπορίου.

3.3 Παραγοντοποίηση πίνακα (Matrix Factorization)

Η παραγοντοποίηση πίνακα (Matrix Factorization, MF) αποτελεί τη σπουδαιότερη τεχνική ανάπτυξης που προέκυψε από τη δεκαετία 2000-2010 στο πεδίο των συστημάτων συστάσεων. Η βασική ιδέα είναι η αποσύνθεση του πίνακα αλληλεπίδρασης χρήστη-αντικειμένου R ∈

R^(|U|×|I|) σε δύο πίνακες χαμηλότερης τάξης: R ≈ P · Q^T, όπου P ∈ R^(|U|×k) είναι ο πίνακας λανθανουσών αναπαραστάσεων χρηστών και Q ∈ R^(|I|×k) ο πίνακας λανθανουσών αναπαραστάσεων αντικειμένων, με k << min(|U|, |I|). Κάθε χρήστης u αναπαρίσταται από το διάνυσμα p_u και κάθε αντικείμενο i από το διάνυσμα q_i, ενώ η προβλεπόμενη βαθμολογία είναι το εσωτερικό τους γινόμενο: r̂(u,i) = p_u^T · q_i.

Η εκπαίδευση των πινάκων P και Q γίνεται με ελαχιστοποίηση της συνάρτησης σφάλματος ανακατασκευής στις γνωστές βαθμολογίες, συνήθως με Stochastic Gradient Descent (SGD) ή Alternating Least Squares (ALS). Ο αλγόριθμος ALS είναι ιδιαίτερα κατάλληλος για κατανεμημένη επεξεργασία, καθώς εναλλάξ σταθεροποιεί τον έναν πίνακα και βελτιστοποιεί τον άλλον, επιλύοντας κάθε φορά ένα γραμμικό σύστημα. Η τεχνική regularization (L2 κανονικοποίηση) εφαρμόζεται για την αποφυγή υπερπροσαρμογής (overfitting).

3.3.1 Βασικά μοντέλα παραγοντοποίησης

Τα κυριότερα μοντέλα MF που έχουν προταθεί στη βιβλιογραφία και χρησιμοποιούνται ευρέως είναι τα ακόλουθα. Η Αποσύνθεση Ιδιαζουσών Τιμών (Singular Value Decomposition, SVD) αποτελεί τη μαθηματικά καθαρότερη μορφή παραγοντοποίησης, ωστόσο στην πράξη ο κλασικός SVD δεν εφαρμόζεται απευθείας σε αραιούς πίνακες. Η Probabilistic Matrix Factorization (PMF) εισήγαγε πιθανοτική ερμηνεία της MF, μοντελοποιώντας τις βαθμολογίες ως Γκαουσιανές τυχαίες μεταβλητές. Το BPR (Bayesian Personalized Ranking) βελτιστοποιεί απευθείας τη σχετική κατάταξη αντικειμένων αντί για την απόλυτη τιμή βαθμολογίας, κάτι που το καθιστά ιδιαίτερα κατάλληλο για σενάρια implicit feedback. Τέλος, το SVD++ επεκτείνει τον κλασικό SVD ενσωματώνοντας πληροφορία από αντικείμενα που ο χρήστης έχει αξιολογήσει έμμεσα.

Αξίζει να τονιστεί ότι ο διαγωνισμός Netflix Prize (2006-2009) έπαιξε καθοριστικό ρόλο στην εδραίωση των τεχνικών MF ως state-of-the-art για πρόβλεψη βαθμολογιών. Οι νικητήριες λύσεις ήταν υβριδικά ensembles που συνδύαζαν πολλαπλά μοντέλα MF, κάτι που ανέδειξε επίσης τη δύναμη των υβριδικών προσεγγίσεων στο πεδίο.

3.3.2 Περιορισμοί της παραγοντοποίησης πίνακα

Παρά την επιτυχία της, η MF αντιμετωπίζει ορισμένους θεμελιώδεις περιορισμούς που ώθησαν στην ανάπτυξη των μεθόδων βαθιάς μάθησης. Ο πιο σημαντικός είναι ότι το εσωτερικό γινόμενο p_u^T · q_i αποτελεί γραμμική συνάρτηση, αδυνατώντας να συλλάβει πολύπλοκες μη γραμμικές σχέσεις μεταξύ χρηστών και αντικειμένων. Όπως επισήμανε ο He et al. (2017), η MF αντιμετωπίζει τις αλληλεπιδράσεις χρήστη-αντικειμένου ως αμιγώς γραμμική πληροφορία, χάνοντας έτσι υποψίες αλληλεπίδρασης που δεν εκφράζονται γραμμικά. Αυτό ακριβώς το κενό ήρθαν να καλύψουν τα νευρωνικά δίκτυα στο επόμενο στάδιο εξέλιξης του πεδίου.

3.4 Αναλυτική σύγκριση μαθηματικών ιδιοτήτων μεθόδων MF

Η επιλογή μεταξύ SVD, PMF, ALS και BPR εξαρτάται από τον τύπο των διαθέσιμων δεδομένων και τον στόχο βελτιστοποίησης. Ο Πίνακας 3.1 συνοψίζει τις μαθηματικές ιδιότητες κάθε μεθόδου, αναδεικνύοντας τους λόγους για τους οποίους το BPR έχει καταστεί η βασική επιλογή για σενάρια implicit feedback, ενώ το ALS παραμένει ανταγωνιστικό για κατανεμημένη επεξεργασία μεγάλης κλίμακας.

Πίνακας 3.1: Μαθηματικές ιδιότητες μεθόδων Matrix Factorization

Μέθοδος	Συνάρτηση Κόστους	Αλγόριθμος	Τύπος Δεδομένων	Πολυπλοκότητα
SVD	$\|\|R - P \cdot Q^T\|\|^2$	Gradient Descent	Ρητή ανάδραση	$O(k \cdot \|R\|)$
PMF	Gaussian likelihood	MAP estimation	Ρητή ανάδραση	$O(k \cdot \|R\|)$
ALS	$\|\|R - P \cdot Q^T\|\|^2 + \lambda \cdot reg$	Alternating LS	Ρητή & Έμμεση	$O(k^2 \cdot \|R\|)$
BPR	Pairwise ranking loss	SGD με triplets	Έμμεση ανάδραση	$O(k \cdot \|DS\|)$
SVD++	$\|\|R - P' \cdot Q^T\|\|^2 + \lambda \cdot reg$	SGD	Ρητή + Έμμεση	$O(k \cdot \|R\| + k \cdot \|N\|)$

3.5 Συστήματα βασισμένα σε γνώση (Knowledge-based systems)

Τα συστήματα βασισμένα σε γνώση (knowledge-based recommender systems) αποτελούν μια διαφορετική κατηγορία που δεν στηρίζεται σε ιστορικά δεδομένα αλληλεπίδρασης, αλλά σε δομημένη γνώση για τον τομέα εφαρμογής. Ο αλγόριθμος χρησιμοποιεί κανόνες, οντολογίες ή γνωσιακούς γράφους για να συνδέσει τις απαιτήσεις του χρήστη με τα χαρακτηριστικά των αντικειμένων, παράγοντας συστάσεις χωρίς να απαιτείται ιστορικό αξιολόγησης.

Η προσέγγιση αυτή είναι ιδιαίτερα αποτελεσματική σε πεδία όπου τα αντικείμενα είναι πολύπλοκα και δύσκολο να αξιολογηθούν απλά (π.χ. ακίνητα, αυτοκίνητα, ιατρικές θεραπείες), όπου ο χρήστης έχει συγκεκριμένες και πολλαπλές απαιτήσεις, και όπου η ασφάλεια δεν επιτρέπει παράλειψη κρίσιμων χαρακτηριστικών. Η βασική αδυναμία των knowledge-based συστημάτων είναι ότι απαιτούν σημαντική ανθρώπινη εργασία για την κατασκευή και συντήρηση της γνωσιακής βάσης, κάτι που δυσχεραίνει την κλιμάκωσή τους σε νέους τομείς.

3.6 Υβριδικά συστήματα (Hybrid systems)

Τα υβριδικά συστήματα συστάσεων συνδυάζουν δύο ή περισσότερες από τις παραπάνω προσεγγίσεις, με στόχο να αξιοποιήσουν τα δυνατά σημεία κάθε μεθόδου ενώ παράλληλα εξουδετερώνουν τις αδυναμίες της. Η βιβλιογραφία αναγνωρίζει διάφορες στρατηγικές υβριδισμού: η σταθμισμένη συνδυαστική προσέγγιση (weighted hybridization) συνδυάζει τα αποτελέσματα πολλών αλγορίθμων με σταθερά ή δυναμικά βάρη, ενώ η διαδοχική υβριδοποίηση (cascade hybridization) χρησιμοποιεί μια μέθοδο ως φίλτρο και μια άλλη για τελική αξιολόγηση.

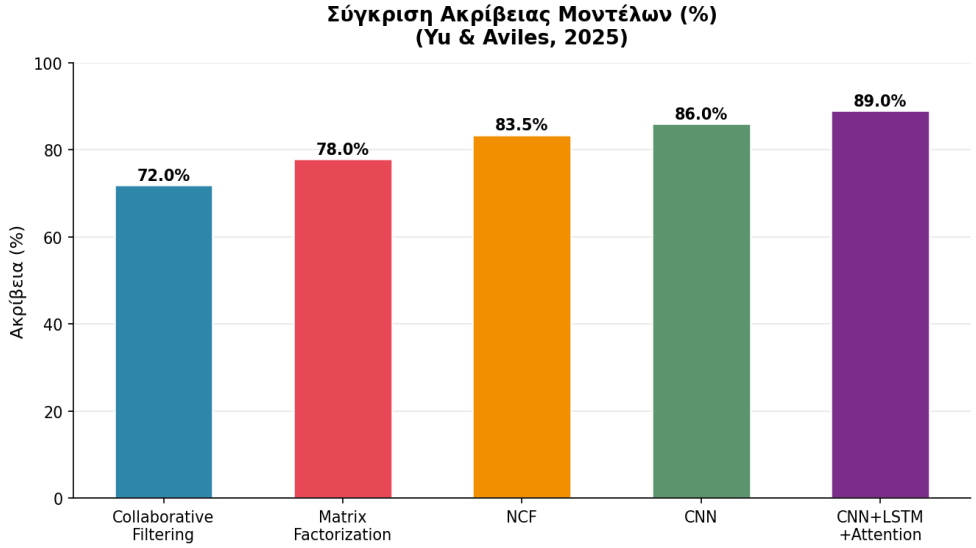
Ένα χαρακτηριστικό παράδειγμα υβριδικής προσέγγισης που εξετάζεται στη βιβλιογραφία είναι η ενσωμάτωση ανάλυσης συναισθήματος (sentiment analysis) στα συστήματα συνεργατικής διήθησης. Οι Darraz et al. (2025) πρότειναν ένα υβριδικό μοντέλο που συνδυάζει τεχνικές λεξιλογικής ανάλυσης με νευρωνικά δίκτυα πολλαπλών επιπέδων (MLP) για την ανάλυση κειμενικών κριτικών χρηστών, και στη συνέχεια ενσωματώνει τα συναισθηματικά σκορ στο μοντέλο DeepMF για βελτιστοποιημένες συστάσεις ξενοδοχείων. Το μοντέλο πέτυχε ακρίβεια ανάλυσης συναισθήματος 88,63% και RMSE 0,1, καταδεικνύοντας την αποτελεσματικότητα της υβριδικής προσέγγισης.

Οι Wang et al. (2025) ανέπτυξαν ένα υβριδικό μοντέλο που συνδυάζει παραγοντοποίηση πίνακα με βαθιά νευρωνικά δίκτυα, επιτυγχάνοντας καλύτερη επίδοση από μεμονωμένα μοντέλα SVD και CF ως προς ακρίβεια, ανάκληση και RMSE. Το αποτέλεσμα αυτό επιβεβαιώνει τη γενικότερη παρατήρηση της βιβλιογραφίας ότι οι υβριδικές μέθοδοι υπερτερούν σταθερά των μεμονωμένων προσεγγίσεων, ιδίως σε πρακτικές εφαρμογές μεγάλης κλίμακας.

Πίνακας 3.2: Συγκριτική αξιολόγηση κλασικών μεθόδων σύστασης

Μέθοδος	Βασική Αρχή	Πλεονεκτήματα	Μειονεκτήματα
Content-based	Ομοιότητα χαρακτηριστικών αντικειμένων	Δεν απαιτεί άλλους χρήστες · Ερμηνεύσιμο	Over-specialization · Cold start νέου χρήστη
User-based CF	Παρόμοιοι χρήστες προτείνουν αντικείμενα	Αξιοποιεί κοινωνική πληροφορία	Δύσκολη κλιμάκωση · Αραιότητα
Item-based CF	Παρόμοια αντικείμενα με ήδη αγαπημένα	Πιο σταθερό από user-based	Cold start νέου αντικειμένου
Matrix Factorization	Λανθάνουσα αναπαράσταση χρηστών/αντικειμένων	Αποτελεσματικό με αραιά δεδομένα	Μόνο γραμμικές σχέσεις · Ψυχρή εκκίνηση
Hybrid	Συνδυασμός δύο ή περισσότερων μεθόδων	Ανθεκτικότητα σε αδυναμίες	Αυξημένη πολυπλοκότητα
Knowledge-based	Γνωσιακοί γράφοι / Κανόνες	Λειτουργεί με λίγα δεδομένα	Απαιτεί χειροκίνητη κατασκευή γνώσης

Το Διάγραμμα 3.1 παρουσιάζει σύγκριση ακρίβειας μεταξύ κλασικών μεθόδων (CF, MF) και σύγχρονων αρχιτεκτονικών βαθιάς μάθησης. Η σταδιακή βελτίωση από 72% (CF) έως 89% (CNN+LSTM+Attention) αναδεικνύει τη σημασία της υιοθέτησης μη γραμμικών μοντέλων. Αξιοσημείωτο είναι ότι ακόμα και η απλή μετάβαση από CF σε MF αποφέρει βελτίωση 6 ποσοστιαίων μονάδων, ενώ η ενσωμάτωση μηχανισμού attention στο CNN+LSTM προσθέτει επιπλέον 3 μονάδες σε σχέση με το απλό CNN.

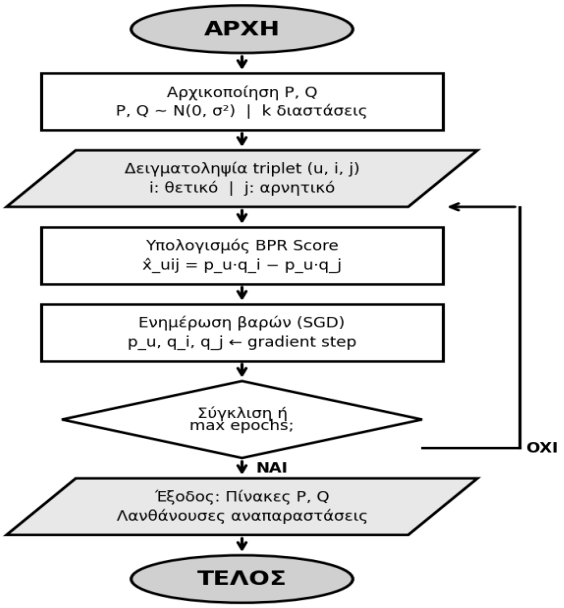


Διάγραμμα 3.1: Σύγκριση ακρίβειας κλασικών μεθόδων και αρχιτεκτονικών DL

3.7 Flowchart αλγορίθμου BPR

Το Σχήμα 3.1 παρουσιάζει τον αλγόριθμο εκπαίδευσης Bayesian Personalized Ranking (BPR) βήμα προς βήμα. Η διαδικασία ξεκινά με την αρχικοποίηση των πινάκων P και Q με τιμές από κανονική κατανομή. Σε κάθε βήμα εκπαίδευσης δειγματοληπτείται ένα triplet (u, i, j), υπολογίζεται το BPR score και ενημερώνονται τα βάρη με SGD. Ο βρόχος τερματίζει όταν ο αλγόριθμος συγκλίνει ή φτάσει στον μέγιστο αριθμό epochs.

Σχήμα 3.1: Εκπαίδευση Matrix Factorization με BPR



Σχήμα 3.1: Αλγόριθμος εκπαίδευσης Matrix Factorization με BPR

3.8 Σύγκριση User-based και Item-based CF: βήμα-βήμα

Η διαφορά μεταξύ user-based και item-based CF γίνεται πλήρως κατανοητή μέσω ενός συγκεκριμένου αριθμητικού παραδείγματος. Χρησιμοποιώντας τον πίνακα αλληλεπίδρασης του Πίνακα 2.3, εξετάζουμε πώς κάθε προσέγγιση θα έκανε τις ίδιες συστάσεις.

Στη user-based CF: (1) Επιλέγουμε τον στοχευόμενο χρήστη u (εδώ η Ελένη). (2) Υπολογίζουμε τις ομοιότητες sim(Ελένη, x) για κάθε άλλο χρήστη x. (3) Επιλέγουμε τους k=2 πλησιέστερους γείτονες: Άλκης (0.71) και Δήμητρα (0.67). (4) Για κάθε αξιολόγηση των γειτόνων σε ταινίες που η Ελένη δεν έχει δει, υπολογίζουμε σταθμισμένο μέσο. (5) Παράγουμε την κατάταξη: Inception (4.0) > Avatar (3.5). Πλεονέκτημα: εξατομίκευση βάσει κοινωνικής ομοιότητας. Μειονέκτημα: αλλάζουν συχνά οι ομοιότητες.

Στη item-based CF: (1) Υπολογίζουμε τις ομοιότητες μεταξύ ταινιών βάσει των κοινών αξιολογήσεών τους. (2) sim(Matrix, Inception) = cos([5, 4, 0, 5, 4], [0, 4, 3, 4, 4]) = (4×4+5×4+4×4)/(√66 × √48) ≈ 0.85. (3) sim(Matrix, Avatar) = cos([5, 4, 0, 5, 4], [3, 0, 0, 4, 0]) = (5×3+5×4)/(√66×√25) ≈ 0.73. (4) Επειδή η Ελένη βαθμολόγησε Matrix=4, οι συστάσεις βασίζονται στις ομοιότητες: Inception (0.85×4=3.4) > Avatar (0.73×4=2.9). Πλεονέκτημα: σταθερές ομοιότητες που υπολογίζονται offline. Μειονέκτημα: λιγότερο εξατομικευμένο.

3.9 Πλήρες αριθμητικό παράδειγμα: SVD βήμα προς βήμα

Για να γίνει περισσότερο κατανοητή η λειτουργία της Παραγοντοποίησης Πίνακα, παρουσιάζεται ένα πλήρες αριθμητικό παράδειγμα εφαρμογής της Singular Value Decomposition (SVD) σε έναν απλό, πλήρως συμπληρωμένο πίνακα αξιολόγησης 4 χρηστών × 4 ταινιών, χωρίς ελλιπείς τιμές, ώστε να είναι εφικτός ο ακριβής υπολογισμός της αποσύνθεσης:

R = [[5,3,1,2], [4,5,2,1], [1,2,5,3], [2,1,3,5]]. Η εφαρμογή SVD στον πίνακα R δίνει R = U·Σ·V^t, όπου οι ιδιάζουσες τιμές (singular values) προκύπτουν ίσες με σ_1 = 11.27, σ_2 = 5.26, σ_3 = 2.79 και σ_4 = 0.74. Παρατηρείται ότι οι δύο πρώτες ιδιάζουσες τιμές είναι σημαντικά μεγαλύτερες από τις υπόλοιπες δύο, υποδεικνύοντας ότι ο πίνακας R έχει ουσιαστικά «ενδογενή διάσταση» περίπου ίση με 2: δηλαδή οι προτιμήσεις των χρηστών εξηγούνται επαρκώς από δύο λανθάνοντες παράγοντες (π.χ. «δράση/περιπέτεια» έναντι «δράμα»).

Κρατώντας μόνο τις δύο πρώτες ιδιάζουσες τιμές (rank-2 προσέγγιση, k=2), υπολογίζεται ο ανακατασκευασμένος πίνακας $\hat{R}$ = U_k·Σ_k·V_k^t, ο οποίος παρουσιάζεται στον Πίνακα 3.3 σε σύγκριση με τις πραγματικές τιμές. Η ποσότητα ενέργειας (energy) που διατηρείται από τις δύο πρώτες ιδιάζουσες τιμές, υπολογιζόμενη ως (σ_1^2 + σ_2^2) / Σσ_i^2, ανέρχεται σε 94.9%, γεγονός που επιβεβαιώνει ότι η rank-2 προσέγγιση αποτυπώνει ικανοποιητικά τη δομή του αρχικού πίνακα, με σχετικό σφάλμα ανακατασκευής (Frobenius) 22.6%.

Πίνακας 3.3: Σύγκριση πραγματικών και εκτιμώμενων τιμών μετά από rank-2 προσέγγιση SVD

Χρήστης	Κατάσταση	Ταινία 1	Ταινία 2	Ταινία 3	Ταινία 4
Χρήστης 1	Πραγματική	5	3	1	2
	Εκτιμώμενη (rank-2)	4.24	3.84	1.51	1.47
Χρήστης 2	Πραγματική	4	5	2	1
	Εκτιμώμενη (rank-2)	4.68	4.24	1.53	1.49
Χρήστης 3	Πραγματική	1	2	5	3
	Εκτιμώμενη (rank-2)	1.49	1.47	4.00	4.01
Χρήστης 4	Πραγματική	2	1	3	5
	Εκτιμώμενη (rank-2)	1.53	1.51	3.99	4.00

Η σύγκριση αναδεικνύει το βασικό πλεονέκτημα της SVD ως τεχνικής συμπίεσης πληροφορίας: ο αρχικός πίνακας 4×4 (16 τιμές) προσεγγίζεται ικανοποιητικά από μόλις δύο λανθάνοντες παράγοντες ανά χρήστη και ταινία (συνολικά 16 παράμετροι U, Σ, V για k=2, αλλά με σημαντικά μικρότερη πραγματική πολυπλοκότητα λόγω της δομής τους). Στην πράξη, όπως αναφέρθηκε στην Ενότητα 3.3, οι πίνακες αξιολόγησης περιέχουν συνήθως πολυάριθμες ελλιπείς τιμές, γεγονός που εμποδίζει τον απευθείας υπολογισμό της κλασικής SVD και οδηγεί στη χρήση επαναληπτικών μεθόδων όπως η ALS ή το Gradient Descent, που προσεγγίζουν το ίδιο αποτέλεσμα χωρίς να απαιτούν πλήρως συμπληρωμένο πίνακα.

3.10 Παραλλαγές ALS: Explicit έναντι Implicit

Ο αλγόριθμος Alternating Least Squares (ALS), που παρουσιάστηκε στην Ενότητα 3.3, διαθέτει δύο βασικές παραλλαγές ανάλογα με τον τύπο ανάδρασης που χειρίζεται. Στην εκδοχή για ρητή ανάδραση (explicit ALS), η συνάρτηση κόστους ελαχιστοποιεί απευθείας το σφάλμα πρόβλεψης μόνο στα παρατηρούμενα ζεύγη χρήστη-αντικειμένου, αγνοώντας πλήρως τα μη παρατηρούμενα ζεύγη, τα οποία θεωρούνται απλώς «άγνωστα» και δεν συμμετέχουν στη βελτιστοποίηση.

Στην εκδοχή για έμμεση ανάδραση (implicit ALS), που εισήγαγαν αρχικά οι Hu, Koren και Volinsky, όλα τα ζεύγη χρήστη-αντικειμένου συμμετέχουν στη βελτιστοποίηση, με τα μη παρατηρούμενα ζεύγη να αντιμετωπίζονται ως ασθενείς αρνητικές ενδείξεις με χαμηλό βάρος εμπιστοσύνης, ενώ τα παρατηρούμενα ζεύγη λαμβάνουν βάρος ανάλογο της συχνότητας ή έντασης της αλληλεπίδρασης (Ενότητα 2.9). Αυτή η θεμελιώδης διαφορά εξηγεί γιατί η implicit ALS κλιμακώνεται διαφορετικά: ενώ η explicit ALS έχει υπολογιστικό κόστος ανάλογο μόνο του αριθμού των παρατηρούμενων αξιολογήσεων |R|, η implicit ALS απαιτεί ειδικές αλγοριθμικές βελτιστοποιήσεις (π.χ. τον αλγόριθμο conjugate gradient) ώστε να αποφευχθεί ο υπολογισμός πάνω σε ολόκληρο τον πυκνό πίνακα |U|×|I|.

3.11 Συνοπτική σύγκριση πλεονεκτημάτων και μειονεκτημάτων

Ο Πίνακας 3.4 συνοψίζει τα κύρια πλεονεκτήματα και μειονεκτήματα των κλασικών μεθόδων σύστασης που αναλύθηκαν στο παρόν κεφάλαιο, παρέχοντας μια συγκριτική επισκόπηση που διευκολύνει την επιλογή κατάλληλης μεθόδου ανάλογα με τα χαρακτηριστικά της εφαρμογής.

Πίνακας 3.4: Συνοπτική σύγκριση πλεονεκτημάτων και μειονεκτημάτων κλασικών μεθόδων

Μέθοδος	Πλεονεκτήματα	Μειονεκτήματα
User-based CF	Απλή υλοποίηση, διαισθητική	Δεν κλιμακώνεται καλά, ευάλωτη σε αραιότητα
Item-based CF	Πιο σταθερή γειτονιά αντικειμένων	Απαιτεί προ-υπολογισμένο πίνακα ομοιότητας
Content-based	Λειτουργεί χωρίς ιστορικό άλλων χρηστών	Περιορισμένη ανακάλυψη (over-specialization)
Matrix Factorization	Καλή απόδοση, χειρίζεται αραιότητα	Δυσκολία στη ψυχρή εκκίνηση
Υβριδικά συστήματα	Συνδυάζει πλεονεκτήματα πολλών μεθόδων	Αυξημένη πολυπλοκότητα σχεδίασης

3.12 Συστάσεις βασισμένες σε γράφο πριν τα GNN: Random Walks

Πριν την εμφάνιση των γραφικών νευρωνικών δικτύων που αναλύονται στο Κεφάλαιο 5, η αναπαράσταση δεδομένων σύστασης ως γράφου χρηστών-αντικειμένων είχε ήδη αξιοποιηθεί μέσω κλασικών τεχνικών τυχαίου περιπάτου (random walk). Αλγόριθμοι όπως ο προσωποποιημένος PageRank (Personalized PageRank) εκτελούν επαναλαμβανόμενους τυχαίους περιπάτους ξεκινώντας από έναν κόμβο-χρήστη, υπολογίζοντας την πιθανότητα επίσκεψης κάθε άλλου κόμβου του γράφου ως μέτρο εγγύτητας (proximity), το οποίο στη συνέχεια χρησιμοποιείται για την κατάταξη πιθανών συστάσεων.

Αυτές οι τεχνικές βάσει γράφου είχαν το πλεονέκτημα ότι μπορούσαν να αποτυπώσουν έμμεσες, πολλαπλών βημάτων σχέσεις μεταξύ χρηστών και αντικειμένων (π.χ. «φίλοι των φίλων» ή «αντικείμενα που αγόρασαν χρήστες με όμοια προφίλ»), κάτι που οι απλές μέθοδοι γειτονιάς της Ενότητας 3.2 δεν μπορούν να αποτυπώσουν απευθείας. Ωστόσο, η απουσία εκμαθημένων αναπαραστάσεων (embeddings) και η αποκλειστική εξάρτηση από τη δομή του γράφου χωρίς ενσωμάτωση χαρακτηριστικών κόμβων αποτέλεσε τον κύριο περιορισμό τους, τον οποίο ήρθαν να καλύψουν οι σύγχρονες αρχιτεκτονικές GNN, συνδυάζοντας τη δομική πληροφορία των τυχαίων περιπάτων με εκμαθημένες αναπαραστάσεις χαρακτηριστικών, όπως αναλύθηκε στο Κεφάλαιο 5.

3.13 Υβριδικές αρχιτεκτονικές: στρατηγικές συνδυασμού

Τα υβριδικά συστήματα σύστασης, που αναφέρθηκαν συνοπτικά στην Ενότητα 3.6, μπορούν να ταξινομηθούν σε διακριτές στρατηγικές συνδυασμού ανάλογα με το πώς συγχωνεύονται οι επιμέρους τεχνικές. Η σταθμισμένη προσέγγιση (weighted hybrid) υπολογίζει ξεχωριστά σκορ από κάθε επιμέρους μέθοδο (π.χ. συνεργατική διήθηση και διήθηση βάσει περιεχομένου) και τα συνδυάζει μέσω γραμμικού ή μη γραμμικού συνδυασμού βαρών. Η εναλλακτική προσέγγιση (switching hybrid) επιλέγει δυναμικά ποια μέθοδο θα χρησιμοποιήσει ανάλογα με τις συνθήκες: για παράδειγμα, διήθηση βάσει περιεχομένου σε σενάρια cold start και συνεργατική διήθηση όταν υπάρχει επαρκές ιστορικό.

Η προσέγγιση χαρακτηριστικής συνδυασμού (feature combination hybrid) ενσωματώνει χαρακτηριστικά από πολλαπλές πηγές ως κοινή είσοδο σε ένα ενιαίο μοντέλο μάθησης, αντί να συνδυάζει ανεξάρτητα σκορ (μια λογική που υιοθετούν φυσικά τα βαθιά μοντέλα του Κεφαλαίου 4. Τέλος, η προσέγγιση καταρράκτη (cascade hybrid) εφαρμόζει τις μεθόδους διαδοχικά, με μια πρώτη, υπολογιστικά ελαφρύτερη μέθοδο να παράγει έναν περιορισμένο κατάλογο υποψηφίων αντικειμένων, και μια δεύτερη, ακριβέστερη αλλά υπολογιστικά βαρύτερη μέθοδο να επανακατατάσσει αυτόν τον κατάλογο) η ίδια αρχιτεκτονική λογική δύο σταδίων που αναλύθηκε στην Ενότητα 9.7 για συστήματα μεγάλης κλίμακας.

3.14 Πρακτικά ζητήματα υλοποίησης κλασικών μεθόδων

Η πρακτική υλοποίηση των κλασικών μεθόδων που αναλύθηκαν στο παρόν κεφάλαιο εγείρει ορισμένα ζητήματα που συχνά παραβλέπονται στη θεωρητική παρουσίαση των αλγορίθμων. Η αποθήκευση του πίνακα αξιολογήσεων ως πυκνού πίνακα (dense matrix) είναι υπολογιστικά απαγορευτική για πραγματικές πλατφόρμες με εκατομμύρια χρήστες και αντικείμενα, καθιστώντας απαραίτητη τη χρήση αραιών αναπαραστάσεων (sparse matrix representations), όπως οι μορφές CSR (Compressed Sparse Row) ή COO (Coordinate List), που αποθηκεύουν μόνο τις μη μηδενικές τιμές.

Επιπλέον, η περιοδική επανεκπαίδευση των μοντέλων παραγοντοποίησης πίνακα σε περιβάλλον παραγωγής, όπου νέες αλληλεπιδράσεις εισέρχονται συνεχώς, απαιτεί στρατηγικές αυξητικής ενημέρωσης (incremental updates) των λανθάνοντων παραγόντων, αντί για πλήρη επανεκπαίδευση από το μηδέν σε κάθε κύκλο ενημέρωσης: μια πρακτική παράμετρος που σχετίζεται άμεσα με το ζήτημα του υπολογιστικού κόστους που αναλύθηκε στην Ενότητα 8.13.

3.15 Σύνοψη

Η επισκόπηση των κλασικών μεθόδων -- διήθηση βάσει περιεχομένου, συνεργατική διήθηση, παραγοντοποίηση πίνακα, υβριδικά συστήματα -- δείχνει ότι καμία δεν είναι απλώς «ξεπερασμένη»: κάθε μία έχει τα δικά της δυνατά και αδύναμα σημεία. Η CBF προσφέρει ερμηνευσιμότητα αλλά πάσχει από υπερεξειδίκευση, η user-based CF αξιοποιεί κοινωνική πληροφορία αλλά δυσκολεύεται στην κλιμάκωση, και η MF επιτυγχάνει εξαιρετικά αποτελέσματα αλλά περιορίζεται από τη γραμμικότητα του μοντέλου της. Η κατανόηση αυτών των ορίων είναι απαραίτητη για να εκτιμηθεί γιατί η εμφάνιση της βαθιάς μάθησης αποτέλεσε τόσο σημαντική τομή στο πεδίο, θέμα που αναλύεται στο επόμενο κεφάλαιο.

4. Deep learning σε συστήματα συστάσεων

Η ενσωμάτωση τεχνικών βαθιάς μάθησης (deep learning) στα συστήματα συστάσεων αποτελεί, κατά τη γνώμη μου, την πιο σημαντική τεχνολογική τομή του πεδίου την τελευταία δεκαετία, ακριβώς επειδή λύνει τον περιορισμό που εντοπίστηκε στο τέλος του προηγούμενου κεφαλαίου: τη γραμμικότητα της MF. Τα νευρωνικά δίκτυα ξεπερνούν αυτόν τον περιορισμό μοντελοποιώντας μη γραμμικές σχέσεις μεταξύ χρηστών και αντικειμένων, μαθαίνοντας αυτόματα ιεραρχικές αναπαραστάσεις από ακατέργαστα δεδομένα και αξιοποιώντας ετερογενείς πηγές πληροφορίας (κείμενο, εικόνα, ακουστικά δεδομένα, γράφοι) σε ένα ενιαίο πλαίσιο -- με κόστος, όπως θα φανεί, σημαντικά μεγαλύτερη υπολογιστική πολυπλοκότητα και μικρότερη ερμηνευσιμότητα.

4.1 Νευρωνική συνεργατική διήθηση (Neural Collaborative Filtering)

Το μοντέλο Neural Collaborative Filtering (NCF), που παρουσιάστηκε από τους He et al. (2017), αποτελεί ένα από τα σημαντικότερα ορόσημα στην ανάπτυξη των νευρωνικών συστημάτων συστάσεων. Η θεμελιώδης συνεισφορά του NCF είναι η αντικατάσταση του γραμμικού εσωτερικού γινομένου της Matrix Factorization με ένα Πολυεπίπεδο Perceptron (Multi-Layer Perceptron, MLP), ικανό να μάθει αυθαίρετες συναρτήσεις από τα δεδομένα εκπαίδευσης. Με αυτό τον τρόπο, το NCF μπορεί να συλλαμβάνει μη γραμμικές και πολύπλοκες αλληλεπιδράσεις χρηστών-αντικειμένων που παραμένουν αόρατες για τη MF.

Η αρχιτεκτονική NeuMF (Neural Matrix Factorization) αποτελεί την πιο ώριμη εκδοχή του NCF, συνδυάζοντας παράλληλα δύο διαδρομές: μια Generalized Matrix Factorization (GMF) που διατηρεί την ικανότητα γραμμικής μοντελοποίησης, και ένα MLP που μαθαίνει μη γραμμικές αλληλεπιδράσεις. Τα αποτελέσματα των δύο διαδρομών συνδυάζονται στο τελικό επίπεδο

πρόβλεψης. Πειράματα σε πραγματικά datasets ταξιδίων έδειξαν ότι το NCF ξεπερνά σημαντικά τις παραδοσιακές μεθόδους CF ως προς ακρίβεια και ικανοποίηση χρήστη, ειδικά σε σενάρια με αραία δεδομένα.

Η αρχιτεκτονική Wide & Deep, που αναπτύχθηκε από την Google για εφαρμογή στο Google Play Store, αποτελεί ένα άλλο σημαντικό μοντέλο της ίδιας οικογένειας. Συνδυάζει μια «πλατιά» γραμμική συνιστώσα για αποστήθιση (memorization) και μια «βαθιά» νευρωνική συνιστώσα για γενίκευση (generalization), επιτυγχάνοντας ισορροπία μεταξύ εξατομίκευσης και κάλυψης.

4.2 Αυτόματα κωδικοποιητικά (Autoencoders)

Τα Autoencoders (AE) είναι νευρωνικά δίκτυα που εκπαιδεύονται να συμπιέζουν (encode) την είσοδο σε μια λανθάνουσα αναπαράσταση χαμηλότερης διάστασης και στη συνέχεια να την ανακατασκευάζουν (decode). Στο πλαίσιο των συστημάτων συστάσεων, ο πίνακας αλληλεπίδρασης χρήστη-αντικειμένου τροφοδοτείται στο δίκτυο, το οποίο μαθαίνει να ανακατασκευάζει τον ίδιο πίνακα (συμπεριλαμβανομένων των ελλειπόντων στοιχείων) ταυτόχρονα μαθαίνοντας μια συμπαγή λανθάνουσα αναπαράσταση των προτιμήσεων.

Το μοντέλο AutoRec εφαρμόζει αυτή την ιδέα απευθείας στον πίνακα αξιολογήσεων, παράγοντας προβλέψεις για ελλείπουσες βαθμολογίες μέσω ανακατασκευής. Τα Collaborative Denoising Autoencoders (CDAE) επεκτείνουν την ιδέα προσθέτοντας τεχνητό θόρυβο στην είσοδο κατά την εκπαίδευση, αναγκάζοντας το δίκτυο να μάθει πιο ανθεκτικές αναπαραστάσεις. Οι Restricted Boltzmann Machines (RBM), ένας τύπος στοχαστικού νευρωνικού δικτύου, εφαρμόστηκαν επίσης με επιτυχία: σε πειράματα στο dataset MovieLens-1M, RBM-based συστήματα πέτυχαν RMSE 0,8736 και MAE 0,6854, σημαντικά καλύτερα από τις αντίστοιχες κλασικές μεθόδους CF.

4.3 Αναδρομικά νευρωνικά δίκτυα (RNN, LSTM, GRU)

Τα Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks, RNNs) σχεδιάστηκαν για την επεξεργασία δεδομένων που έχουν φυσική διαδοχική ή χρονική δομή. Στο πλαίσιο των συστημάτων συστάσεων, η ακολουθία των αντικειμένων με τα οποία έχει αλληλεπιδράσει ένας χρήστης στο χρόνο αποτελεί φυσική «πρόταση» που τα RNNs μπορούν να επεξεργαστούν για να προβλέψουν το επόμενο αντικείμενο ενδιαφέροντος. Η ικανότητα των RNNs να «θυμούνται» πληροφορία από προηγούμενα βήματα μέσω της κρυφής κατάστασης (hidden state) τα καθιστά ιδανικά για session-based recommendation, όπου ο στόχος είναι η πρόβλεψη της επόμενης αλληλεπίδρασης εντός μιας σύντομης περιόδου χρήσης.

4.3.1 LSTM και GRU

Τα Long Short-Term Memory (LSTM) δίκτυα, μια εξελιγμένη παραλλαγή των RNN, εισήγαγαν τον μηχανισμό «πύλης» (gating mechanism) για την αντιμετώπιση του προβλήματος της εξαφανιζόμενης βαθμίδας (vanishing gradient), που εμποδίζει τα απλά RNNs από το να μαθαίνουν μακροπρόθεσμες εξαρτήσεις. Τα LSTM διατηρούν μια «κυψέλη μνήμης» (cell state) που μπορεί να αποθηκεύει πληροφορία για μεγάλα χρονικά διαστήματα, ελεγχόμενη από τρεις πύλες: εισόδου, λήθης και εξόδου. Τα Gated Recurrent Units (GRU) αποτελούν μια απλοποιημένη παραλλαγή των LSTM με λιγότερες παραμέτρους, συχνά συγκρίσιμης ή καλύτερης απόδοσης σε πρακτικές εφαρμογές.

Το μοντέλο GRU4Rec αποτελεί ένα από τα πρώτα και πιο επιδραστικά μοντέλα session-based σύστασης βασισμένα σε GRU. Το Hierarchical RNN (HRNN) επεκτείνει την ιδέα μοντελοποιώντας τόσο ενδο-session (within-session) όσο και δια-session (cross-session) εξαρτήσεις σε ιεραρχική δομή. Οι Yoon & Jang (2023) επισημαίνουν ότι τα RNN-based μοντέλα, παρά την επιτυχία τους, υστερούν στη σύλληψη μακροπρόθεσμων εξαρτήσεων λόγω της σειριακής τους φύσης, πρόβλημα που ήρθαν να λύσουν τα Transformer-based μοντέλα.

4.4 Νευρωνικά δίκτυα συνέλιξης (CNN)

Τα Νευρωνικά Δίκτυα Συνέλιξης (Convolutional Neural Networks, CNNs) αρχικά σχεδιάστηκαν για επεξεργασία εικόνων, αλλά εφαρμόστηκαν με επιτυχία και στα συστήματα συστάσεων, κυρίως για την εξαγωγή χαρακτηριστικών από κείμενα κριτικών χρηστών. Τα φίλτρα

συνέλιξης ενός CNN λειτουργούν ως ανιχνευτές τοπικών προτύπων (local pattern detectors): εφαρμόζονται διαδοχικά σε τμήματα της εισόδου, εξάγοντας n-gram χαρακτηριστικά που αντιπροσωπεύουν τοπικές σχέσεις μεταξύ λέξεων ή στοιχείων.

Ένα χαρακτηριστικό παράδειγμα αξιοποίησης CNN στα συστήματα συστάσεων είναι το μοντέλο ADRS (Adaptive Deep learning-based Recommendation System) των Da'u et al. (2021). Το ADRS ενσωματώνει ένα attentive CNN για την εκμάθηση προσαρμοστικών αναπαραστάσεων χρηστών και αντικειμένων από κειμενικές κριτικές. Σε αντίθεση με τα στατικά μοντέλα που χρησιμοποιούν σταθερά διανύσματα χαρακτηριστικών, το ADRS εφαρμόζει έναν μηχανισμό mutual attention που συλλαμβάνει λεπτομερείς αλληλεπιδράσεις χρήστη-αντικειμένου, επιτυγχάνοντας σημαντικά καλύτερα αποτελέσματα σε datasets Amazon και Yelp σε σχέση με τα baseline μοντέλα.

Το μοντέλο Caser χρησιμοποιεί οριζόντια και κάθετα φίλτρα συνέλιξης για να συλλαμβάνει τόσο point-level όσο και union-level σχέσεις μεταξύ διαδοχικών αντικειμένων, μοντελοποιώντας αποτελεσματικά σύντομα σειριακά πρότυπα. Το ConvNCF εφαρμόζει εξωτερικό γινόμενο μεταξύ embeddings χρήστη και αντικειμένου, δημιουργώντας έναν δισδιάστατο «χάρτη αλληλεπίδρασης» τον οποίο επεξεργάζονται φίλτρα CNN, αξιοποιώντας τόσο τοπικές όσο και ολικές σχέσεις.

4.5 Transformers και μηχανισμοί προσοχής (Attention mechanisms)

Η αρχιτεκτονική Transformer, που παρουσιάστηκε το 2017 από τους Vaswani et al. για την επεξεργασία φυσικής γλώσσας, επανάστασε επίσης το πεδίο των συστημάτων συστάσεων. Η καινοτομία του Transformer έγκειται στον μηχανισμό self-attention, που επιτρέπει σε κάθε στοιχείο μιας ακολουθίας να «προσέχει» άμεσα οποιοδήποτε άλλο στοιχείο, ανεξαρτήτως απόστασης. Αυτό λύνει το πρόβλημα των μακροπρόθεσμων εξαρτήσεων που παρουσιάζουν τα RNNs, ενώ επιτρέπει ταυτόχρονη (παράλληλη) επεξεργασία όλων των στοιχείων: σε αντίθεση με τη σειριακή φύση των RNN.

4.5.1 Self-Attention Sequential Recommendation

Το μοντέλο SASRec (Self-Attention based Sequential Recommendation) αποτελεί ένα από τα πρώτα και πιο επιδραστικά Transformer-based μοντέλα για sequential recommendation. Χρησιμοποιεί μονόδρομο (unidirectional) self-attention για να μοντελοποιήσει την ακολουθία αλληλεπιδράσεων ενός χρήστη, προβλέποντας το επόμενο αντικείμενο ενδιαφέροντος. Η αρχιτεκτονική του επιτρέπει τη σύλληψη τόσο βραχυπρόθεσμων όσο και μακροπρόθεσμων εξαρτήσεων στο ιστορικό του χρήστη.

Το BERT4Rec εφαρμόζει την αρχιτεκτονική BERT (Bidirectional Encoder Representations from Transformers) στα συστήματα συστάσεων, χρησιμοποιώντας αμφίδρομο self-attention και masked item prediction ως στόχο εκπαίδευσης. Αυτή η αμφίδρομη φύση επιτρέπει βαθύτερη κατανόηση των σχέσεων μεταξύ αντικειμένων σε σχέση με τα μονόδρομα μοντέλα. Οι Zangari et al. (2026) επισημαίνουν ότι οι Transformer-based αρχιτεκτονικές, όταν συνδυάζονται με γραφικές δομές, συχνά υπερτερούν των GNN σε benchmarks σύστασης, καθώς αποφεύγουν το φαινόμενο over-smoothing που παρουσιάζουν τα GNN.

4.6 Γεννητικά ανταγωνιστικά δίκτυα (GAN)

Τα Generative Adversarial Networks (GANs) εισήχθησαν στο πεδίο των συστημάτων συστάσεων ως εργαλείο βελτίωσης της ποιότητας και ποικιλομορφίας των συστάσεων. Η αρχιτεκτονική ενός GAN αποτελείται από δύο νευρωνικά δίκτυα: έναν Γεννήτη (Generator, G) που παράγει

«πλαστές» συστάσεις ή δεδομένα, και έναν Διακριτή (Discriminator, D) που προσπαθεί να διακρίνει τα «αληθινά» δεδομένα από τα παραγόμενα. Τα δύο δίκτυα εκπαιδεύονται σε ανταγωνιστικό παίγνιο (minimax game), με αποτέλεσμα ο Generator να βελτιώνεται σταδιακά στην παραγωγή ρεαλιστικών δεδομένων.

Στα συστήματα συστάσεων, τα GANs εφαρμόζονται κυρίως για δύο σκοπούς. Πρώτον, για τη δημιουργία συνθετικών δεδομένων εκπαίδευσης (data augmentation), αντιμετωπίζοντας έτσι το πρόβλημα αραιότητας. Δεύτερον, για τη βελτίωση της ποικιλομορφίας των συστάσεων, καθώς ο Generator μαθαίνει να παράγει συστάσεις που δεν περιορίζονται στα πιο δημοφιλή αντικείμενα. Χαρακτηριστικά μοντέλα είναι το IRGAN, που χρησιμοποιεί GAN για βελτιστοποίηση της κατάταξης (ranking), και το CFGAN, που εφαρμόζει GAN για collaborative filtering.

4.7 Αυτο-επιβλεπόμενη μάθηση (Self-Supervised Learning)

Η Αυτο-επιβλεπόμενη Μάθηση (Self-Supervised Learning, SSL) αποτελεί μια από τις πιο σύγχρονες και ταχύτατα αναπτυσσόμενες κατευθύνσεις στο πεδίο. Η βασική ιδέα είναι η δημιουργία σημάτων επίβλεψης (supervision signals) απευθείας από τα δεδομένα, χωρίς ανάγκη χειροκίνητης σήμανσης (labeling). Αυτό επιτυγχάνεται μέσω τεχνικών επαύξησης δεδομένων (data augmentation) που δημιουργούν «θετικά ζεύγη» (positive pairs) αντικειμένων ή χρηστών, τα οποία ορίζονται ως παρόμοια και πρέπει να έχουν κοντινές αναπαραστάσεις στον λανθάνοντα χώρο (latent space).

Η αντιπαραθετική μάθηση (Contrastive Learning) είναι η πιο διαδεδομένη μορφή SSL στα συστήματα συστάσεων. Μοντέλα όπως το SGL (Self-supervised Graph Learning) και το SimGCL εφαρμόζουν contrastive learning σε Graph Neural Networks για συστάσεις, επιτυγχάνοντας σημαντικές βελτιώσεις έναντι των αντίστοιχων μοντέλων χωρίς SSL. Το CL4Rec εφαρμόζει contrastive learning σε sequential recommendation, βελτιώνοντας την ποιότητα των αναπαραστάσεων ακόμα και με αραιά δεδομένα. Η SSL προσέγγιση είναι ιδιαίτερα πολύτιμη γιατί αντιμετωπίζει ταυτόχρονα δύο κρίσιμα προβλήματα: την data sparsity και τη μεροληψία δημοτικότητας.

Πίνακας 4.1: Συγκριτική επισκόπηση αρχιτεκτονικών βαθιάς μάθησης στα συστήματα συστάσεων

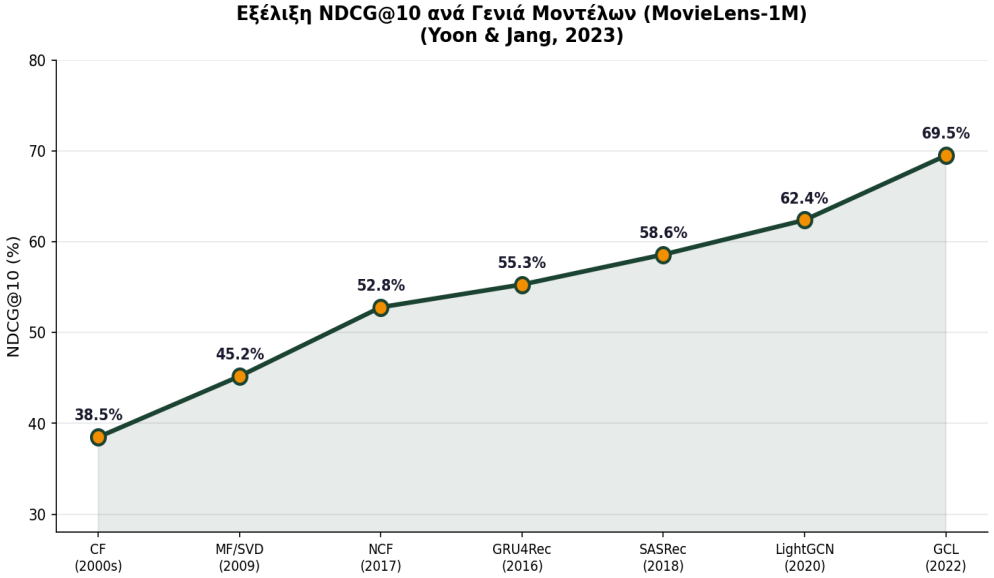
Αρχιτεκτονική	Βασική Αρχή	Εφαρμογή στις Συστάσεις	Βασικά Μοντέλα
MLP / NCF	Μη γραμμική μοντελοποίηση χρήστη-αντικειμένου	Αντικατάσταση εσωτερικού γινομένου MF	NeuMF, Wide & Deep
Autoencoder	Κωδικοποίηση και ανακατασκευή δεδομένων	Πρόβλεψη ελλειπουσών βαθμολογιών	AutoRec, CDAE, RBM
RNN / LSTM / GRU	Σειριακή επεξεργασία χρονικά διαδοχικών δεδομένων	Session-based & Sequential Recommendation	GRU4Rec, HRNN, SASRec
CNN	Τοπική εξαγωγή χαρακτηριστικών με φίλτρα συνέλιξης	Μοντελοποίηση κειμένου κριτικών, εικόνων	ADRS, Caser, ConvNCF
Transformer / Attention	Μηχανισμός self-attention για μακροπρόθεσμες εξαρτήσεις	Σειριακή σύσταση, Graph-based RS	BERT4Rec, SASRec, BSARec
GAN	Γεννήτρια vs Διακριτής σε ανταγωνιστική εκπαίδευση	Βελτίωση ποικιλομορφίας, augmentation δεδομένων	IRGAN, CFGAN, RecGAN
SSL / Contrastive	Αυτο-επιβλεπόμενη μάθηση χωρίς ετικέτες	Αντιμετώπιση αραιότητας, βελτίωση embeddings	SGL, SimGCL, CL4Rec

4.8 Αναπαραστάσεις διανυσμάτων (Embeddings)

Ένα κοινό θεμέλιο σε σχεδόν όλες τις αρχιτεκτονικές βαθιάς μάθησης για συστάσεις είναι η χρήση embeddings: πυκνών διανυσματικών αναπαραστάσεων χρηστών και αντικειμένων σε έναν συνεχή λανθάνοντα χώρο (latent space). Σε αντίθεση με τις αραιές one-hot αναπαραστάσεις, τα embeddings μαθαίνονται κατά τη διάρκεια της εκπαίδευσης και κωδικοποιούν σημασιολογικές ομοιότητες: αντικείμενα με παρόμοια χαρακτηριστικά τείνουν να έχουν κοντινά διανύσματα στον λανθάνοντα χώρο. Η ποιότητα των embeddings επηρεάζει άμεσα την απόδοση του συστήματος, γι' αυτό πολλές σύγχρονες έρευνες εστιάζουν στη βελτιστοποίησή τους.

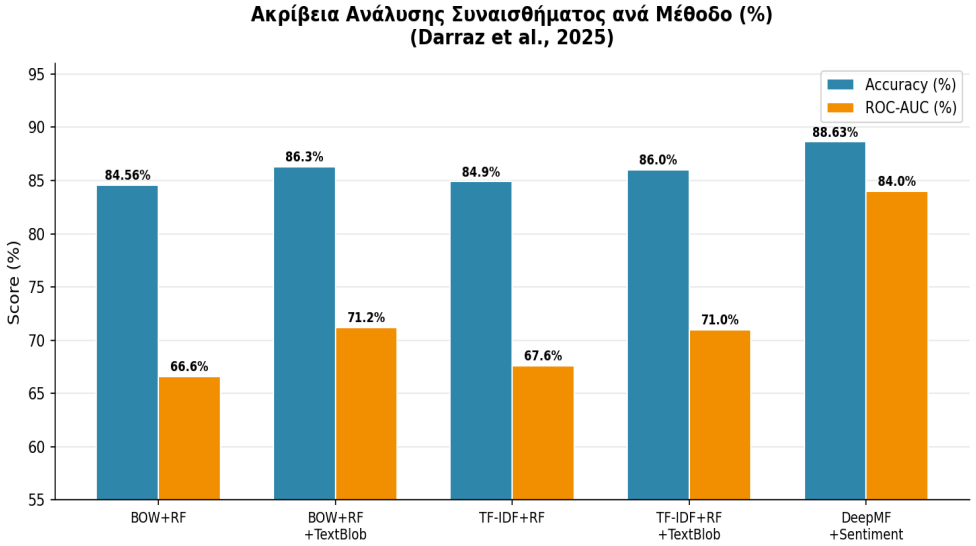
Τεχνικές όπως το Word2Vec και το Item2Vec έδειξαν ότι η ακολουθιακή εκπαίδευση embeddings αντικειμένων (αντιμετωπίζοντας ακολουθίες αντικειμένων ως «προτάσεις» σε αναλογία με λέξεις σε κείμενο) παράγει εμπλουτισμένες αναπαραστάσεις που συλλαμβάνουν σχέσεις συμπληρωματικότητας και υποκατάστασης. Πιο πρόσφατα, η τεχνική Graph Embedding αξιοποιεί τη δομή του γράφου αλληλεπίδρασης χρήστη-αντικειμένου για την παραγωγή ακόμα πλουσιότερων αναπαραστάσεων, ανοίγοντας τον δρόμο για τα Γραφικά Νευρωνικά Δίκτυα που αναλύονται στο επόμενο κεφάλαιο.

Το Διάγραμμα 4.1 αποτυπώνει χρονολογικά την εξέλιξη της μετρικής NDCG@10 στο MovieLens-1M, ανά γενιά μοντέλων. Από τα απλά CF (38.5%) έως τα GCL μοντέλα (69.5%) παρατηρείται αύξηση σχεδόν 31 ποσοστιαίων μονάδων. Το μεγαλύτερο μεμονωμένο άλμα εμφανίζεται μεταξύ MF/SVD (45.2%) και NCF (52.8%), καταδεικνύοντας την αξία της αντικατάστασης του γραμμικού εσωτερικού γινομένου με MLP. Επίσης σημαντική είναι η μετάβαση από LightGCN (62.4%) σε GCL (69.5%), που οφείλεται στην αυτο-επιβλεπόμενη εκπαίδευση.



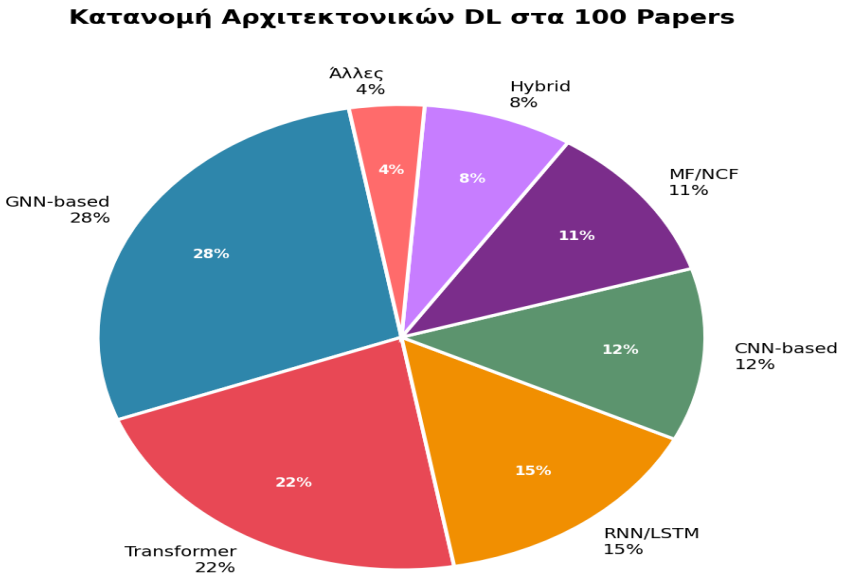
Διάγραμμα 4.1: Εξέλιξη NDCG@10 ανά γενιά μοντέλων στο MovieLens-1M

Το Διάγραμμα 4.2 συγκρίνει μεθόδους ανάλυσης συναισθήματος για sentiment-enhanced σύσταση. Η ενσωμάτωση TextBlob βελτιώνει συστηματικά τόσο την ακρίβεια όσο και το ROC-AUC σε κάθε μέθοδο. Το υβριδικό μοντέλο DeepMF+Sentiment επιτυγχάνει 88.63% ακρίβεια και 84% ROC-AUC, σημαντικά ανώτερα από τα baseline μοντέλα BOW+RF (84.56% / 66.60%) και TF-IDF+RF. Η βελτίωση ROC-AUC από 67.6% σε 84.0% αντιπροσωπεύει αύξηση 16.4 ποσοστιαίων μονάδων: ενδεικτική της προστιθέμενης αξίας της βαθιάς μάθησης για ανάλυση συναισθήματος.



Διάγραμμα 4.2: Σύγκριση μεθόδων ανάλυσης συναισθήματος για RS

Το Διάγραμμα 4.3 δείχνει την κατανομή αρχιτεκτονικών βαθιάς μάθησης στα 100 papers της βιβλιογραφικής ανασκόπησης. Τα GNN-based μοντέλα κυριαρχούν με 28%, αντικατοπτρίζοντας τη στροφή της έρευνας προς γραφικές αναπαραστάσεις. Τα Transformer-based μοντέλα (22%) βρίσκονται σε ταχεία άνοδο, ενώ τα RNN/LSTM (15%) αρχίζουν να υποχωρούν. Τα υβριδικά μοντέλα (8%) αναδεικνύουν την τάση συνδυασμού πολλαπλών τεχνικών.



Διάγραμμα 4.3: Κατανομή αρχιτεκτονικών βαθιάς μάθησης στα 100 εξεταζόμενα papers

4.9 Ψυχολογική μοντελοποίηση καταναλωτή με GNN

Μια αναδυόμενη κατεύθυνση στο πεδίο της βαθιάς μάθησης για συστάσεις είναι η ενσωμάτωση ψυχολογικών μοντέλων καταναλωτή. Οι παραδοσιακές μέθοδοι (CNN, RNN, MF) αντιμετωπίζουν τη συμπεριφορά χρήστη ως ακολουθία ενεργειών, αγνοώντας τους ψυχολογικούς παράγοντες που τις κινούν: παρορμητισμό, εμπιστοσύνη, συναισθηματική κατάσταση, προσοχή. Ο Gao (2026) πρότεινε το NeuroGraph-CPM (Consumer Psychological Modeling), ένα μοντέλο που συνδυάζει GNN με ψυχολογικά σήματα από κριτικές και αλληλεπιδράσεις χρήστη.

Η αρχιτεκτονική NeuroGraph-CPM κατασκευάζει έναν ετερογενή γράφο χρήστη-προϊόντος που περιλαμβάνει δεδομένα συμπεριφοράς, περιβαλλοντικές πληροφορίες και ψυχολογικά σήματα από κριτικές. Ο μηχανισμός affect-aware message passing επιτρέπει στο δίκτυο να αναπαριστά τόσο σταθερές (stable) όσο και δυναμικές (dynamic) γνωσιακές-συναισθηματικές σχέσεις. Πειράματα στο Amazon Electronics dataset έδειξαν βελτίωση ακρίβειας κατά 19.6%, CTR estimation κατά 16.3% και personalization relevance κατά 21.8% σε σχέση με ισχυρά baseline μοντέλα.

Το αξιοσημείωτο χαρακτηριστικό του NeuroGraph-CPM είναι η ερμηνευσιμότητά του: τα βάρη προσοχής (attention weights) συνδέονται με συγκεκριμένα ψυχολογικά prior, επιτρέποντας την εξήγηση του γιατί ένα αντικείμενο συστήνεται βάσει ψυχολογικών αντί μόνο συμπεριφορικών χαρακτηριστικών. Αυτό αποτελεί σημαντικό πλεονέκτημα για εφαρμογές marketing, όπου η κατανόηση των κινήτρων αγοράς είναι εξίσου σημαντική με την ακρίβεια σύστασης.

Πίνακας 4.2: Σύγκριση μοντέλων ψυχολογικής μοντελοποίησης καταναλωτή

Μοντέλο	Βελτίωση Ακρίβειας	Βελτίωση CTR	Βασική Καινοτομία
CNN baseline	—	—	Τυπικό CNN χωρίς ψυχολογικά χαρακτηριστικά
RNN baseline	—	—	Σειριακή μοντελοποίηση χωρίς context
PANE-GNN	+8.3%	+6.1%	Graph + static psychological features
NeuroGraph-CPM	+19.6%	+16.3%	Affect-aware message passing + psychological regularization

4.10 Εξερεύνηση και μετρικές αξιολόγησης βασισμένες σε Bayesian Information Gain

Ένα θεμελιώδες δίλημμα των συστημάτων συστάσεων είναι η ισορροπία μεταξύ εκμετάλλευσης (exploitation) (πρόταση αντικειμένων υψηλής βεβαιότητας προτίμησης) και εξερεύνησης (exploration): πρόταση αντικειμένων αβέβαιης αλλά δυνητικά ενδιαφέρουσας προτίμησης. Αν ένα σύστημα βασίζεται αποκλειστικά σε εκμετάλλευση, ο χρήστης εκτίθεται σε περιορισμένο εύρος αντικειμένων, αναπτύσσεται φυσαλίδα φίλτρων και η μακροπρόθεσμη ικανοποίηση υποβαθμίζεται. Αν αντίθετα η εξερεύνηση είναι υπερβολική, η βραχυπρόθεσμη εμπειρία χρήστη υποβαθμίζεται λόγω χαμηλής άμεσης σχετικότητας.

Οι Ishii et al. (2026) πρότειναν το BIG-PU (Bayesian Information Gain for Preference Updates), μια νέα μετρική αξιολόγησης εξερεύνησης από τη σκοπιά του συστήματος. Η BIG-PU μετρά κατά πόσο η σύσταση βελτιώνει την κατανόηση του συστήματος για τις προτιμήσεις του χρήστη, αξιοποιώντας την αβεβαιότητα (uncertainty) των εκτιμούμενων κατανομών προτίμησης. Τυπικά, ορίζεται ως το κέρδος πληροφορίας (information gain) από την αλληλεπίδραση χρήστη-αντικειμένου: BIG-PU(u, i) = H(θ_u) - E[H(θ_u | interaction(u,i))], όπου H() είναι η εντροπία της κατανομής προτίμησης χρήστη u και θ_u οι παράμετροι αυτής.

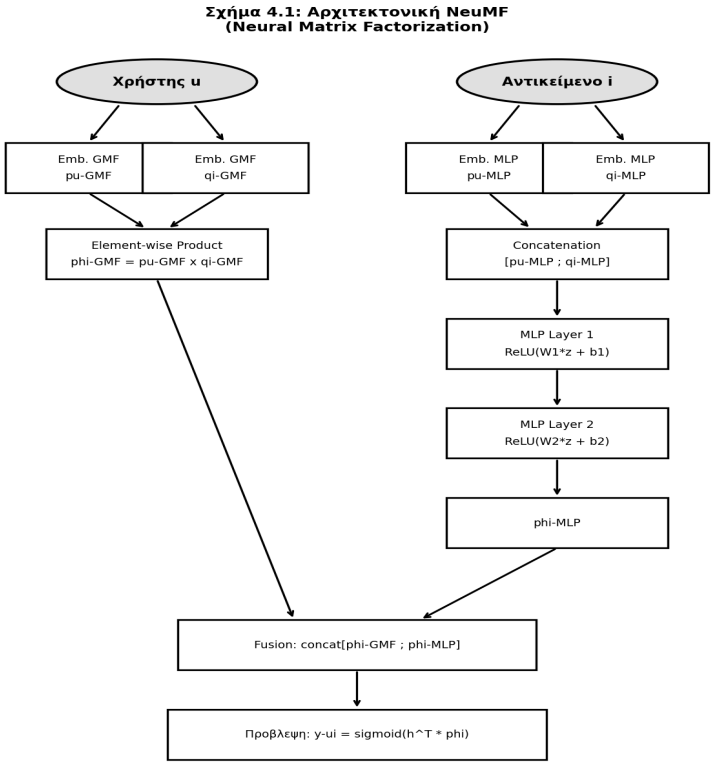
Η BIG-PU υπερτερεί έναντι των υπαρχουσών μετρικών (Diversity, Novelty, Serendipity) σε δύο βασικά σημεία. Πρώτον, είναι πιο γενική: ενώ η Diversity εξαρτάται από τον ορισμό ομοιότητας αντικειμένων και η Novelty από στατιστικά δημοτικότητας, η BIG-PU μετρά απευθείας τη βελτίωση της κατανόησης του χρήστη ανεξαρτήτως αυτών. Δεύτερον, είναι σύμφωνη με μακροπρόθεσμα οφέλη: υψηλή BIG-PU σημαίνει ότι η σύσταση προσφέρει νέα πληροφορία, κάτι που βελτιώνει τις μελλοντικές συστάσεις. Πειράματα σε συνθετικά και πραγματικά datasets επιβεβαίωσαν την αποτελεσματικότητά της.

Πίνακας 4.3: Σύγκριση μετρικών αξιολόγησης εξερεύνησης

Μετρική Εξερεύνησης	Τύπος	Βασικό Μειονέκτημα	Περιγραφή
Diversity	Item-based	Εξαρτάται από ορισμό	Μετρά ποικιλία αντικειμένων στη λίστα
Novelty	Popularity-based	Αγνοεί προτιμήσεις	Προτείνει λιγότερο γνωστά αντικείμενα
Serendipity	Relevance+Surprise	Δύσκολο να μετρηθεί	Ευχάριστη έκπληξη στις συστάσεις
BIG-PU	Bayesian Info Gain	Υψηλότερη πολυπλοκότητα	Μετρά βελτίωση κατανόησης προτίμησης χρήστη

4.11 Αρχιτεκτονικό διάγραμμα NeuMF

Το Σχήμα 4.1 απεικονίζει αναλυτικά την αρχιτεκτονική NeuMF (Neural Matrix Factorization). Η είσοδος χρήστη και αντικειμένου τροφοδοτείται παράλληλα σε δύο διαδρομές: την GMF (Generalized Matrix Factorization) που υπολογίζει element-wise γινόμενο των embeddings, και την MLP που εφαρμόζει διαδοχικά ReLU layers για σύλληψη μη γραμμικών αλληλεπιδράσεων. Τα αποτελέσματα των δύο διαδρομών συγχωνεύονται και τροφοδοτούνται στον τελικό νευρώνα εξόδου με sigmoid ενεργοποίηση.



Σχήμα 4.1: Αρχιτεκτονική NeuMF: GMF και MLP σε παράλληλες διαδρομές

4.12 Μαθηματική ανάλυση μηχανισμού Self-Attention

Ο μηχανισμός self-attention αποτελεί τον πυρήνα της αρχιτεκτονικής Transformer και η κατανόησή του σε μαθηματικό επίπεδο είναι απαραίτητη για την αξιολόγηση των Transformer-based συστημάτων συστάσεων. Δεδομένης μιας ακολουθίας εισόδου X ∈ R^(n×d), όπου n είναι το μήκος της ακολουθίας και d η διάσταση των embeddings, ο μηχανισμός self-attention μετασχηματίζει αυτή την ακολουθία σε τρεις αναπαραστάσεις: Query Q, Key K και Value V, μέσω εκπαιδεύσιμων πινάκων W_Q, W_K, W_V ∈ R^(d×d_k).

Η εξίσωση του scaled dot-product attention είναι: Attention(Q, K, V) = softmax(Q·K^T / √d_k) · V, όπου ο διαιρέτης √d_k εισάγεται για σταθεροποίηση των κλίσεων κατά την εκπαίδευση: χωρίς αυτόν, τα εσωτερικά γινόμενα Q·K^T τείνουν να γίνονται πολύ μεγάλα σε υψηλές διαστάσεις, οδηγώντας σε εξαιρετικά μικρές κλίσεις της softmax. Το αποτέλεσμα της softmax αντιπροσωπεύει τα βάρη προσοχής (attention weights): κάθε στοιχείο της ακολουθίας «προσέχει» τα άλλα στοιχεία με διαφορετική ένταση, εξάγοντας σχετική πληροφορία.

Στο Multi-Head Attention (MHA), ο μηχανισμός επαναλαμβάνεται h φορές παράλληλα με διαφορετικούς πίνακες βαρών, επιτρέποντας το μοντέλο να «βλέπει» την ακολουθία από h διαφορετικές οπτικές γωνίες ταυτόχρονα: MH(Q, K, V) = concat(head_1, ..., head_h) · W_O, όπου head_i = Attention(Q·W_Q^i, K·W_K^i, V·W_V^i). Στα συστήματα συστάσεων, αυτό επιτρέπει το μοντέλο να συλλαμβάνει ταυτόχρονα βραχυπρόθεσμες και μακροπρόθεσμες εξαρτήσεις στην ακολουθία αλληλεπιδράσεων.

Πίνακας 4.4: Μαθηματικές συνιστώσες του μηχανισμού Transformer

Συνιστώσα	Μαθηματικός Ορισμός	Ρόλος στα RS
Query (Q)	Q = X·W_Q, W_Q ∈ R^(d×d_k)	Τι αναζητεί ο χρήστης
Key (K)	K = X·W_K, W_K ∈ R^(d×d_k)	Ποιο αντικείμενο ταιριάζει
Value (V)	V = X·W_V, W_V ∈ R^(d×d_v)	Τι πληροφορία μεταφέρεται
Attention	Att(Q,K,V) = softmax(QK^T/√d_k)·V	Σταθμισμένη συνάθροιση
Multi-Head	MH(Q,K,V) = concat(head_1,...,head_h)·W_O	Πολλαπλές οπτικές γωνίες
Positional Enc.	PE(pos,2i) = sin(pos/10000^(2i/d))	Κωδικοποίηση θέσης ακολουθίας

Η πολυπλοκότητα του self-attention είναι O(n²·d), καθιστώντας τον υπολογιστικά απαιτητικό για μακρές ακολουθίες. Αυτό αποτελεί σημαντικό περιορισμό σε σενάρια sequential recommendation με μακρά ιστορικά. Για την αντιμετώπισή του, έχουν προταθεί προσεγγιστικές παραλλαγές όπως το Linformer και το Performer, που μειώνουν την πολυπλοκότητα σε O(n·d) μέσω τεχνικών random feature approximation ή low-rank decomposition.

4.13 Variational Autoencoders (VAE) για συστάσεις

Πέραν των βασικών αυτόματων κωδικοποιητών (autoencoders) που παρουσιάστηκαν στην Ενότητα 4.2, οι Variational Autoencoders (VAE) αποτελούν μια πιθανοτική επέκτασή τους που έχει αποδειχθεί ιδιαίτερα αποτελεσματική σε εφαρμογές σύστασης. Αντί να κωδικοποιούν την είσοδο σε ένα σταθερό διάνυσμα λανθάνοντος χώρου, οι VAE μαθαίνουν μια κατανομή πιθανότητας q_φ(z|x) πάνω από τον latent space (συνήθως μια πολυμεταβλητή κανονική κατανομή N(μ, σ²)) επιτρέποντας στο μοντέλο να αποτυπώσει την αβεβαιότητα γύρω από τις προτιμήσεις του χρήστη, αντί για μια απλή σημειακή εκτίμηση.

Η εκπαίδευση των VAE βασίζεται στη μεγιστοποίηση ενός κάτω φράγματος της πιθανοφάνειας (Evidence Lower Bound, ELBO): ELBO = E_q_φ(z|x)[ln p_θ(x|z)] - D_KL(q_φ(z|x) ‖ p(z)). Ο πρώτος όρος εκφράζει την ικανότητα ανακατασκευής (reconstruction) των αλληλεπιδράσεων του χρήστη, ενώ ο δεύτερος όρος, η απόκλιση Kullback-Leibler, λειτουργεί ως regularization, εξαναγκάζοντας την κατανομή q_φ(z|x) να προσεγγίζει την εκ των

προτέρων κατανομή p(z), συνήθως μια τυπική κανονική N(0, I). Το τέχνασμα της επαναπαραμετροποίησης (reparameterization trick), $z = \mu + \sigma \odot \varepsilon$ με $\varepsilon \sim N(0,I)$, επιτρέπει τη διάδοση της κλίσης μέσα από τη στοχαστική δειγματοληψία, καθιστώντας δυνατή την εκπαίδευση με τυπικές μεθόδους οπισθοδιάδοσης.

Σε σχέση με τους ντετερμινιστικούς autoencoders, οι VAE προσφέρουν δύο πλεονεκτήματα στο πλαίσιο των συστημάτων συστάσεων: πρώτον, ο λανθάνων χώρος είναι πιο ομαλός (smooth) και συνεχής, επιτρέποντας πιο σταθερή γενίκευση σε χρήστες με αραιά δεδομένα. Δεύτερον, η στοχαστική φύση της κωδικοποίησης λειτουργεί ως εγγενής μηχανισμός κανονικοποίησης (regularization), μειώνοντας την υπερπροσαρμογή (overfitting) σε σχέση με τους κλασικούς αυτόματους κωδικοποιητές, ιδιαίτερα σε σύνολα δεδομένων με υψηλό βαθμό αραιότητας, όπως αναλύθηκε στην Ενότητα 2.6.

4.14 Αναπαραστάσεις αντικειμένων με Item2Vec και BERT4Rec

Εμπνευσμένο από το μοντέλο word2vec της επεξεργασίας φυσικής γλώσσας, το Item2Vec αντιμετωπίζει κάθε ακολουθία αλληλεπιδράσεων ενός χρήστη ως «πρόταση», όπου τα αντικείμενα παίζουν τον ρόλο λέξεων, και εκπαιδεύει embeddings αντικειμένων ώστε αντικείμενα που εμφανίζονται συχνά μαζί στο ίδιο πλαίσιο (context window) να αποκτούν κοντινές αναπαραστάσεις στον latent space. Η προσέγγιση αυτή αγνοεί τη σειρά των αλληλεπιδράσεων, εστιάζοντας αποκλειστικά στη συν-εμφάνιση (co-occurrence), και αποτέλεσε πρόδρομο των σύγχρονων αρχιτεκτονικών σειριακής σύστασης που παρουσιάστηκαν στην Ενότητα 4.6.

Το BERT4Rec επεκτείνει αυτή την ιδέα αξιοποιώντας πλήρως αμφίδρομο (bidirectional) self-attention, σε αντίθεση με το μονόδρομο SASRec που αναλύθηκε στην Ενότητα 4.5. Κατά την εκπαίδευση, το BERT4Rec χρησιμοποιεί έναν στόχο τύπου «masked language modeling»: τυχαία αντικείμενα της ακολουθίας αλληλεπιδράσεων καλύπτονται (masked) και το μοντέλο καλείται να τα προβλέψει χρησιμοποιώντας πληροφορία από αμφότερες τις κατευθύνσεις της ακολουθίας. Αυτή η αμφίδρομη κωδικοποίηση επιτρέπει στο μοντέλο να αξιοποιήσει πλουσιότερο πλαίσιο σε σχέση με τις μονόδρομες προσεγγίσεις, με αντάλλαγμα μεγαλύτερη υπολογιστική πολυπλοκότητα κατά την εκπαίδευση.

4.15 Τεχνικές κανονικοποίησης (regularization) σε βαθιά μοντέλα σύστασης

Η υπερπροσαρμογή (overfitting) αποτελεί σημαντικό κίνδυνο στα βαθιά μοντέλα σύστασης, ιδιαίτερα όταν ο αριθμός των εκπαιδευόμενων παραμέτρων είναι μεγάλος σε σχέση με τον όγκο των διαθέσιμων δεδομένων αλληλεπίδρασης. Η τεχνική dropout, που απενεργοποιεί τυχαία ένα ποσοστό νευρώνων σε κάθε εποχή εκπαίδευσης, εφαρμόζεται ευρέως στα κρυφά επίπεδα των MLP κλάδων αρχιτεκτονικών όπως το NeuMF (Ενότητα 4.1), εξαναγκάζοντας το δίκτυο να μην βασίζεται υπερβολικά σε συγκεκριμένους νευρώνες ή χαρακτηριστικά.

Η regularization L2 (weight decay) προστίθεται στη συνάρτηση κόστους ως ποινή ανάλογη του τετραγώνου των παραμέτρων του μοντέλου, περιορίζοντας το μέγεθος των εκμαθημένων βαρών και εμμέσως την πολυπλοκότητα του μοντέλου, κατά τα πρότυπα της κανονικοποίησης που εφαρμόζεται και στη συνάρτηση κόστους BPR (Παράρτημα Γ, Ενότητα Γ.1). Η κανονικοποίηση παρτίδας (batch normalization), τέλος, σταθεροποιεί την κατανομή των ενδιάμεσων αναπαραστάσεων σε κάθε επίπεδο, επιταχύνοντας τη σύγκλιση και επιτρέποντας τη χρήση μεγαλύτερων ρυθμών μάθησης, αν και η εφαρμογή της σε αρχιτεκτονικές GNN απαιτεί ιδιαίτερη προσοχή λόγω της μεταβλητής δομής του γράφου εισόδου.

4.16 Group Deep Neural Networks και κοινωνικές συστάσεις

Σε αρκετές πραγματικές εφαρμογές, η σύσταση δεν απευθύνεται σε μεμονωμένο χρήστη αλλά σε ομάδα χρηστών: για παράδειγμα, σύσταση ταινίας για μια οικογένεια ή playlist για μια κοινωνική εκδήλωση. Οι αρχιτεκτονικές group deep neural network επεκτείνουν τα βαθιά μοντέλα σύστασης ώστε να συναθροίζουν τις ατομικές προτιμήσεις πολλών χρηστών σε μια ενιαία αναπαράσταση ομάδας, συνήθως μέσω μηχανισμών προσοχής (attention) που μαθαίνουν δυναμικά βάρη επιρροής για κάθε μέλος της ομάδας, αντί για απλή άθροιση ή μέσο όρο των ατομικών προφίλ.

Η προσέγγιση αυτή είναι ιδιαίτερα σημαντική σε δίκτυα κοινωνικής σύστασης (social recommendation), όπου το γράφημα κοινωνικών σχέσεων μεταξύ χρηστών παρέχει πρόσθετη πληροφορία πέραν του γράφου αλληλεπιδράσεων χρήστη-αντικειμένου που αναλύθηκε στο

Κεφάλαιο 5. Η υπόθεση της κοινωνικής επίδρασης (social influence), σύμφωνα με την οποία οι προτιμήσεις ενός χρήστη επηρεάζονται από τις προτιμήσεις των κοινωνικών του επαφών, μπορεί να ενσωματωθεί απευθείας σε αρχιτεκτονικές GNN μέσω ενός δευτερεύοντος γράφου κοινωνικών σχέσεων που συνδυάζεται με τον κύριο γράφο αλληλεπιδράσεων, σε μια μορφή ετερογενούς γράφου παρόμοια με αυτή που περιγράφηκε στην Ενότητα 5.5.

4.17 Υβριδικά μοντέλα Matrix Factorization και Deep Learning

Πέραν των καθαρά νευρωνικών αρχιτεκτονικών που παρουσιάστηκαν στις προηγούμενες ενότητες, μια σημαντική γραμμή έρευνας εστιάζει σε υβριδικούς συνδυασμούς κλασικής παραγοντοποίησης πίνακα (Κεφάλαιο 3) με βαθιά νευρωνικά δίκτυα, με στόχο τον συνδυασμό της ερμηνευσιμότητας και της υπολογιστικής αποδοτικότητας της πρώτης με την εκφραστική ικανότητα της δεύτερης. Σε αυτά τα μοντέλα, οι γραμμικοί λανθάνοντες παράγοντες της MF χρησιμεύουν συχνά ως αρχικοποίηση ή ως συμπληρωματικό σήμα προς ένα βαθύτερο νευρωνικό δίκτυο, αντί να αντικαθίστανται πλήρως από αυτό.

Η στρατηγική αυτή έχει αποδειχθεί ιδιαίτερα αποτελεσματική σε σενάρια όπου ο όγκος δεδομένων δεν είναι αρκετά μεγάλος ώστε να δικαιολογεί την πλήρη πολυπλοκότητα μιας αμιγώς βαθιάς αρχιτεκτονικής, επιτυγχάνοντας καλύτερη γενίκευση σε σχέση με αμιγώς νευρωνικά μοντέλα όπως το NeuMF (Ενότητα 4.1) σε συνθήκες περιορισμένων δεδομένων, ενώ ταυτόχρονα διατηρεί μέρος της ερμηνευσιμότητας των γραμμικών λανθάνοντων παραγόντων: ένα πρακτικό συμβιβασμό ανάμεσα στο trade-off ακρίβειας-ερμηνευσιμότητας που συζητήθηκε στην Ενότητα 9.11.

4.18 Μεταφορά μάθησης (Transfer Learning) σε συστήματα σύστασης

Η μεταφορά μάθησης επιτρέπει την αξιοποίηση γνώσης που αποκτήθηκε σε έναν τομέα ή σύνολο δεδομένων προς όφελος ενός άλλου, συγγενικού τομέα με περιορισμένα δεδομένα: μια τεχνική ιδιαίτερα χρήσιμη για την αντιμετώπιση της cold start συστήματος (Ενότητα 2.11). Στο πλαίσιο της διατομεακής σύστασης (cross-domain recommendation), ένα μοντέλο εκπαιδευμένο σε δεδομένα από μια καθιερωμένη πλατφόρμα (π.χ. ηλεκτρονικό εμπόριο βιβλίων) μπορεί να μεταφέρει εκμαθημένες αναπαραστάσεις χρηστών προς μια νέα πλατφόρμα (π.χ. ηλεκτρονικό εμπόριο ένδυσης), υπό την προϋπόθεση ύπαρξης κοινών χρηστών ή συσχετίσιμων χαρακτηριστικών μεταξύ των δύο τομέων.

Τεχνικά, η μεταφορά μάθησης (transfer learning) σε αρχιτεκτονικές embedding υλοποιείται συχνά μέσω κοινών επιπέδων κωδικοποίησης (shared encoder layers) μεταξύ τομέων, με εξειδικευμένα επιπεδα ανά τομέα (domain-specific layers) στα τελικά στάδια του δικτύου, σε μια λογική ανάλογη της αρχιτεκτονικής δύο κλάδων του NeuMF (Ενότητα 4.1). Η επιτυχία της μεταφοράς μάθησης εξαρτάται κρίσιμα από τον βαθμό συσχέτισης μεταξύ πηγαίου και στόχου τομέα, με υπερβολικά ανόμοιους τομείς να οδηγούν ενδεχομένως σε αρνητική μεταφορά (negative transfer), μειώνοντας αντί να βελτιώνουν την απόδοση στον τομέα-στόχο.

4.19 Σύνοψη

Η εξέταση αυτών των αρχιτεκτονικών αναδεικνύει ένα μοτίβο που θεωρώ σημαντικότερο από την απλή χρονολογική τους σειρά: κάθε νέα προσέγγιση δεν αντικαθιστά την προηγούμενη, αλλά λύνει έναν συγκεκριμένο περιορισμό της με κόστος έναν καινούργιο -- τα NCF κέρδισαν εκφραστικότητα έναντι της MF χάνοντας την ερμηνευσιμότητα του εσωτερικού γινομένου, τα Transformers έλυσαν τις μακροπρόθεσμες εξαρτήσεις των RNN/LSTM με σημαντικά μεγαλύτερο υπολογιστικό κόστος. Η αυτο-επιβλεπόμενη μάθηση φαίνεται να είναι η πιο πρόσφατη απάντηση σε αυτό το δίλημμα, καθώς αντιμετωπίζει την data sparsity χωρίς να απαιτεί επιπλέον ετικέτες. Το ίδιο trade-off επανέρχεται και στα Γραφικά Νευρωνικά Δίκτυα που αναλύονται στο επόμενο κεφάλαιο, τα οποία αξιοποιούν τη δομή γράφου των αλληλεπιδράσεων για την περαιτέρω βελτίωση της ποιότητας των συστάσεων.

5. Graph neural networks για συστάσεις

Τα Γραφικά Νευρωνικά Δίκτυα (Graph Neural Networks, GNNs) αποτελούν σήμερα μια από τις πιο δυναμικά αναπτυσσόμενες κατευθύνσεις στο πεδίο των συστημάτων συστάσεων. Η κεντρική τους ιδέα είναι απλή αλλά ισχυρή: αντί να αντιμετωπίζουν τις αλληλεπιδράσεις χρηστών-αντικειμένων ως έναν επίπεδο πίνακα, τα GNNs μοντελοποιούν αυτές τις

αλληλεπιδράσεις ως ένα γράφο, αξιοποιώντας τη δομή και τη συνδεσιμότητά του για να μάθουν πλουσιότερες αναπαραστάσεις χρηστών και αντικειμένων. Αυτή η προσέγγιση επιτρέπει τη σύλληψη υψηλής τάξης σχέσεων (δηλαδή σχέσεων που δεν είναι άμεσες αλλά υπολανθάνουν μέσα στη δομή του γράφου) που οι παραδοσιακές μέθοδοι αδυνατούσαν να εντοπίσουν.

Ο Zhai (2025) επισημαίνει ότι σε σενάρια σύστασης προϊόντων, χρήστες, αντικείμενα και χαρακτηριστικά τους μπορούν να αναπαρασταθούν φυσικά ως κόμβοι διαφόρων τύπων, ενώ οι αλληλεπιδράσεις μεταξύ τους (αγορές, κλικ, κοινωνικές σχέσεις) αποτελούν τις ακμές του γράφου. Μέσα σε αυτή την αναπαράσταση, τα GNNs εκτελούν επαναληπτικές λειτουργίες μεταβίβασης μηνυμάτων (message passing) που επιτρέπουν σε κάθε κόμβο να συλλέγει πληροφορία από τους γείτονές του και να ενημερώνει τη δική του αναπαράσταση. Μετά από k επαναλήψεις, κάθε κόμβος έχει ενσωματώσει πληροφορία από γείτονες k βημάτων μακριά.

5.1 Βασικές αρχές των γραφικών νευρωνικών δικτύων

Ένα Γραφικό Νευρωνικό Δίκτυο λειτουργεί σε έναν γράφο G = (V, E), όπου V είναι το σύνολο κόμβων και E το σύνολο ακμών. Κάθε κόμβος v ∈ V συνδέεται με ένα αρχικό διάνυσμα χαρακτηριστικών h_v^(0), που τυπικά είναι το embedding χρήστη ή αντικειμένου. Η βασική λειτουργία ενός GNN σε κάθε επίπεδο l είναι: (1) η Συνάθροιση (Aggregation), όπου κάθε κόμβος v συλλέγει τις αναπαραστάσεις των γειτόνων του N(v), και (2) η Ενημέρωση (Update), όπου η αναπαράσταση του v ενημερώνεται συνδυάζοντας την τρέχουσα αναπαράστασή του με την συναθροισμένη πληροφορία. Μαθηματικά: h_v^(l+1) = UPDATE(h_v^(l), AGGREGATE({h_u^(l) : u ∈ N(v)})). Αυτή η διαδικασία επαναλαμβάνεται για L επίπεδα, επιτρέποντας τη σύλληψη πληροφορίας από γείτονες L βημάτων.

Στο πλαίσιο των συστημάτων συστάσεων, ο γράφος χρήστη-αντικειμένου ορίζεται ως ένας δίγραφος (bipartite graph) G = (U ∪ I, A), όπου U είναι το σύνολο χρηστών, I το σύνολο αντικειμένων, και ο πίνακας γειτνίασης A ∈ R^(|U|+|I|) × (|U|+|I|) κωδικοποιεί τις αλληλεπιδράσεις. Ο πίνακας αυτός, ο οποίος αποτελεί επέκταση του κλασικού πίνακα αλληλεπίδρασης χρήστη-αντικειμένου, χρησιμοποιείται ως δομή εισόδου για τις λειτουργίες μεταβίβασης μηνυμάτων. Αυτή η μοντελοποίηση επιτρέπει τη φυσική σύλληψη εξαρτήσεων υψηλής τάξης που αδυνατεί να εντοπίσει η κλασική Matrix Factorization.

5.2 Από το NGCF στο LightGCN: η εξέλιξη του GCN για συστάσεις

Το Neural Graph Collaborative Filtering (NGCF) αποτέλεσε ένα από τα πρώτα και πιο επιδραστικά μοντέλα GNN για συστάσεις. Εισήγαγε τον μηχανισμό embedding propagation, όπου τα embeddings χρηστών και αντικειμένων διαδίδονται διαδοχικά στο γράφο αλληλεπίδρασης. Σε κάθε επίπεδο, κάθε κόμβος ενημερώνει την αναπαράστασή του αθροίζοντας weighted embeddings γειτόνων, με τα βάρη να καθορίζονται από τον κανονικοποιητή συμμετρικού γράφου.

Παρά την επιτυχία του NGCF, οι He et al. (2020) παρατήρησαν ότι πολλές από τις λειτουργίες του (μη γραμμικές ενεργοποιήσεις και μετασχηματισμοί χαρακτηριστικών) ήταν περιττές και αύξαναν την πολυπλοκότητα χωρίς αντίστοιχο κέρδος. Αυτή η παρατήρηση οδήγησε στην ανάπτυξη του LightGCN, που αφαίρεσε τις μη γραμμικές ενεργοποιήσεις και τους μετασχηματισμούς χαρακτηριστικών, διατηρώντας μόνο την απλή σταθμισμένη συνάθροιση γειτόνων. Η τελική αναπαράσταση κάθε κόμβου υπολογίζεται ως μέσος όρος των αναπαραστάσεων από όλα τα επίπεδα. Παρά (ή εξαιτίας) της απλότητάς του, το LightGCN ξεπερνά το NGCF σε ακρίβεια, γεγονός που επιβεβαιώνει εμπειρικά ότι η γραμμική διάδοση πληροφορίας είναι επαρκής για την εκμάθηση αναπαραστάσεων από γράφους αλληλεπίδρασης.

Ο Zhai (2025) εφάρμοσε GNN-based μοντέλα για σύσταση προϊόντων σε περιβάλλον κοινωνικού ηλεκτρονικού εμπορίου (social e-commerce), κατασκευάζοντας έναν ετερογενή γράφο με κόμβους χρηστών, αντικειμένων και χαρακτηριστικών τους. Η χρήση graph convolution και node embedding επέτρεψε την αξιοποίηση πολλαπλών τύπων κόμβων και ακμών, με αποτέλεσμα σημαντική υπεροχή ως προς ακρίβεια και ανάκληση σε σχέση με τους κλασικούς αλγορίθμους CF και content-based.

5.3 Γραφικά δίκτυα προσοχής (Graph Attention Networks)

Τα Graph Attention Networks (GAT) εισήγαγαν στη λειτουργία μεταβίβασης μηνυμάτων έναν μηχανισμό attention, επιτρέποντας σε κάθε κόμβο να αποδίδει διαφορετική σημασία σε διαφορετικούς γείτονες αντί για ομοιόμορφα βάρη. Αυτό αντιμετωπίζει ένα σημαντικό μειονέκτημα των κλασικών GCN: στο πλαίσιο των συστημάτων συστάσεων, δεν έχουν όλες οι

αλληλεπιδράσεις ενός χρήστη την ίδια σχετικότητα με τις τρέχουσες προτιμήσεις του. Ένας χρήστης που αγόρασε κάποτε ένα βιβλίο δώρο για κάποιον άλλο δεν σημαίνει απαραίτητα ότι ενδιαφέρεται για βιβλία.

Στο πλαίσιο των συστημάτων συστάσεων, τα GAT υπολογίζουν βάρη προσοχής (attention weights) e(u, v) μεταξύ κάθε ζεύγους κόμβων (u, v) βάσει των αναπαραστάσεών τους, και στη συνέχεια κανονικοποιούν με softmax για να εξασφαλίσουν συγκρίσιμα βάρη σε διαφορετικές γειτονιές. Αυτός ο μηχανισμός επιτρέπει επίσης την ερμηνευσιμότητα: τα βάρη προσοχής υποδεικνύουν ποιες αλληλεπιδράσεις θεωρεί πιο σημαντικές το μοντέλο, παρέχοντας μια μορφή εξήγησης για τις συστάσεις που παράγει.

5.4 GNN σε γνωσιακούς γράφους (Knowledge Graph-enhanced RS)

Οι Γνωσιακοί Γράφοι (Knowledge Graphs, KGs) αποτελούν μεγάλες δομημένες βάσεις γνώσης που κωδικοποιούν οντότητες και τις σχέσεις μεταξύ τους υπό τη μορφή τριπλέτων (κεφαλή, σχέση, ουρά). Στο πεδίο των συστημάτων συστάσεων, η ενσωμάτωση KG εμπλουτίζει τη σημασιολογική αναπαράσταση των αντικειμένων: ένα αντικείμενο «ταινία» συνδέεται στον KG με τον σκηνοθέτη της, τους ηθοποιούς, το είδος, τη χώρα παραγωγής κ.λπ., παρέχοντας πλούσιο πλαίσιο που τα απλά embeddings αδυνατούν να συλλάβουν.

Μοντέλα όπως το KGCN (Knowledge Graph Convolutional Network) και το KGAT (Knowledge Graph Attention Network) συνδυάζουν τη δύναμη των GNN με τις KG για να μαθαίνουν εμπλουτισμένες αναπαραστάσεις αντικειμένων. Το KGCN εκτελεί επαναληπτική διάδοση πληροφορίας στον KG, επιτρέποντας στα αντικείμενα να «κληρονομούν» χαρακτηριστικά από οντότητες που συνδέονται με αυτά. Το KGAT επεκτείνει αυτή την ιδέα προσθέτοντας attention mechanisms που σταθμίζουν διαφορετικά τους τύπους σχέσεων, επιτρέποντας μεγαλύτερη εκλεπτυσμένη κατανόηση της σημασιολογίας των αντικειμένων.

5.5 Ετερογενείς γράφοι και υπεργράφοι

Στις περισσότερες πραγματικές εφαρμογές, τα δεδομένα δεν αποτελούνται μόνο από ένα τύπο κόμβου και ακμής. Οι Ετερογενείς Γράφοι Συστάσεων (Heterogeneous Graph-based RS) μοντελοποιούν ρητά πολλαπλούς τύπους κόμβων (π.χ. χρήστες, αντικείμενα, κατηγορίες, εμπορικά σήματα) και πολλαπλούς τύπους ακμών (π.χ. αγορά, κλικ, κοινωνική σχέση, ανήκει-σε). Αυτή η πλούσια αναπαράσταση επιτρέπει τη σύλληψη πολύπλοκων σχέσεων που υπολανθάνουν στα δεδομένα πραγματικού κόσμου.

Οι Mall & Singh (2026) παρουσίασαν το HDHP (Hypergraph-Driven Heterogeneous Preference model), ένα μοντέλο που αξιοποιεί υπεργράφους για τη σύλληψη υψηλής τάξης κοινωνικών σχέσεων σε δίκτυα κοινωνικής σύστασης. Σε αντίθεση με τους κανονικούς γράφους, όπου κάθε ακμή συνδέει ακριβώς δύο κόμβους, ένα hyperedge μπορεί να συνδέει οποιοδήποτε αριθμό κόμβων: π.χ. μια ομάδα χρηστών που έχουν αγοράσει το ίδιο προϊόν. Αυτό επιτρέπει τη φυσική μοντελοποίηση σχέσεων υψηλής τάξης που οι δυαδικοί γράφοι αδυνατούν να εκφράσουν. Το HDHP χρησιμοποιεί Hypergraph Convolutional Network (HGCN) για να συναθροίζει πληροφορία από hyperedges, επιτυγχάνοντας recall@20 88,76% και βελτίωση 83,95% σε σχέση με το DHCF baseline στο MovieLens 100K.

5.6 Αντιπαραθετική μάθηση σε γράφους (Graph Contrastive Learning)

Το Graph Contrastive Learning (GCL) αντιπροσωπεύει τη συνέργεια δύο ισχυρών τάσεων: των GNN για γραφική αναπαράσταση και της αυτο-επιβλεπόμενης αντιπαραθετικής μάθησης για βελτίωση των embeddings χωρίς ετικέτες. Η βασική ιδέα είναι η δημιουργία «θετικών ζευγών» κόμβων μέσω τεχνικών επαύξησης γράφου (graph augmentation), όπως η τυχαία αφαίρεση ακμών (edge dropout), η αφαίρεση χαρακτηριστικών κόμβων (node feature masking) ή η δειγματοληψία υπογράφων (subgraph sampling). Δύο views του ίδιου κόμβου που προκύπτουν από διαφορετικές επαυξήσεις αποτελούν θετικό ζεύγος και ο αντίστοιχος κόμβος σε άλλα γράφο views αποτελεί αρνητικό ζεύγος.

Ο Ramteke et al. (2025) εκτέλεσαν εκτεταμένα πειράματα σε MovieLens 100K συγκρίνοντας GCN, GAT, LightGCN και Graph Contrastive Learning. Τα αποτελέσματα έδειξαν ότι το GCL υπερτερεί σε τρεις διαστάσεις: ακρίβεια (precision), ποικιλομορφία (diversity) και δικαιοσύνη (fairness). Ειδικά για χρήστες cold-start και long-tail αντικείμενα (όπου τα παραδοσιακά μοντέλα αποτυγχάνουν λόγω έλλειψης δεδομένων) το GCL επιτυγχάνει σημαντικά καλύτερα αποτελέσματα, χάρη στις πλούσιες αναπαραστάσεις που προκύπτουν από την

αντιπαραθετική εκπαίδευση. Αξίζει να σημειωθεί ότι το GCL δεν απαιτεί πρόσθετα ετικετιαρισμένα δεδομένα: χρησιμοποιεί τα υπάρχοντα αλληλεπιδράσεις δεδομένα για να δημιουργήσει τα δικά του σήματα επίβλεψης.

5.7 GNN για ποικιλομορφία συστάσεων: το MycGNN

Μια ιδιαίτερα καινοτόμα εφαρμογή GNN για τη βελτίωση της ποικιλομορφίας των συστάσεων είναι το MycGNN (Mycelium-inspired Graph Neural Network) των Bahi et al. (2026). Εμπνευσμένο από τη βιολογική δομή του μυκηλίου (του υπόγειου δικτύου νηματίων που σχηματίζουν οι μύκητες και χαρακτηρίζεται από προσαρμοστικότητα, ανθεκτικότητα και ικανότητα εξερεύνησης νέων πόρων) το MycGNN εισάγει δύο βασικούς μηχανισμούς: δυναμική επαναβάρυνση ακμών (dynamic edge reweighting) και μια συνάρτηση εξερεύνησης-εκμετάλλευσης (exploration-exploitation function).

Ο μηχανισμός dynamic edge reweighting τροποποιεί δυναμικά τα βάρη των ακμών μετά την εκπαίδευση του GNN, βάσει των αλληλεπιδράσεων χρηστών και των χαρακτηριστικών αντικειμένων, με σκοπό τη μείωση της μεροληψίας δημοτικότητας. Ο μηχανισμός exploration-exploitation ρυθμίζει τη σχέση μεταξύ της πρότασης οικείων αντικειμένων (εκμετάλλευση) και νέων, μη δημοφιλών αντικειμένων (εξερεύνηση), σε αναλογία με τον τρόπο που το μυκήλιο αξιοποιεί γνωστές πηγές και παράλληλα εξερευνά νέα εδάφη. Πειράματα στα datasets Retail Rocket και MovieLens 1M έδειξαν ότι το MycGNN, αν και δεν υπερτερεί πάντα σε ακρίβεια, επιτυγχάνει σημαντικά υψηλότερη intra-list diversity, καταδεικνύοντας τη δυνατότητα βιο-εμπνευσμένων GNN στην αντιμετώπιση του filter bubble.

5.8 Περιορισμοί των GNN και η στροφή στους Transformers

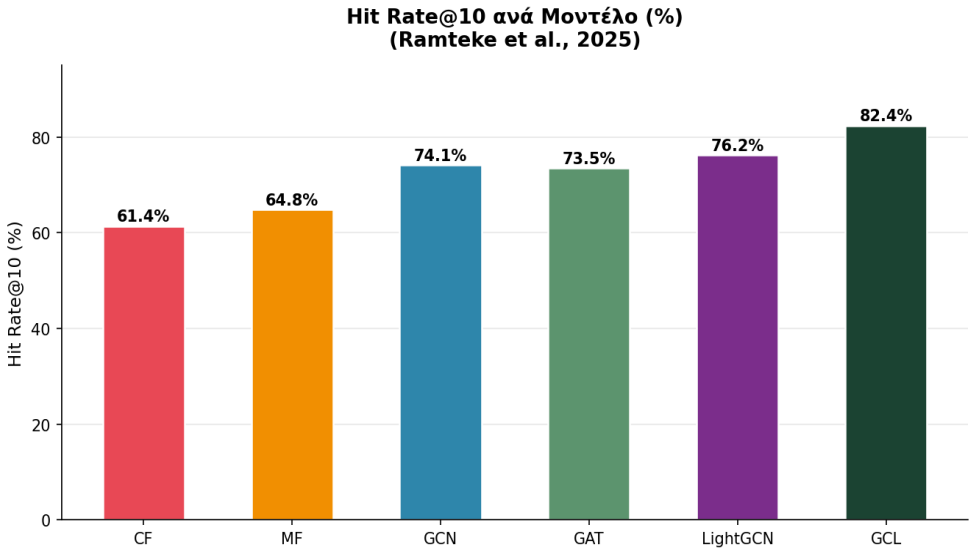
Παρά τις εντυπωσιακές επιδόσεις τους, τα GNN αντιμετωπίζουν ορισμένους θεμελιώδεις περιορισμούς. Ο πιο γνωστός είναι το φαινόμενο της υπεραπλείωσης (over-smoothing): καθώς αυξάνεται ο αριθμός των επιπέδων του GNN, τα embeddings διαφορετικών κόμβων τείνουν να γίνονται ολοένα πιο παρόμοια, χάνοντας τη διακριτικότητα που είναι αναγκαία για ακριβείς συστάσεις. Αυτό περιορίζει πρακτικά τον αριθμό των επιπέδων που μπορούν να χρησιμοποιηθούν, και κατ' επέκταση τη βάθος γειτονιάς που μπορεί να «δει» κάθε κόμβος.

Ένας δεύτερος περιορισμός είναι η επιρρέπεια του μηχανισμού message passing σε θόρυβο και μεροληψία: αν ένας χρήστης έχει αλληλεπιδράσει με αντικείμενα που δεν αντικατοπτρίζουν τις πραγματικές του προτιμήσεις (π.χ. λόγω τυχαίας αγοράς ή επίδρασης δημοτικότητας) αυτές οι άσχετες αλληλεπιδράσεις διαδίδονται στο γράφο και «μολύνουν» τις αναπαραστάσεις. Τρίτον, τα GNN αντιμετωπίζουν δυσκολίες στη σύλληψη μακροπρόθεσμων εξαρτήσεων μεταξύ χρηστών και αντικειμένων λόγω της τοπικής φύσης του message passing. Αυτοί οι περιορισμοί οδήγησαν στη στροφή της έρευνας προς τους Transformers για γράφους (Graph Transformers), που αξιοποιούν το self-attention για να συλλάβουν παγκόσμιες εξαρτήσεις χωρίς τους περιορισμούς του message passing.

Πίνακας 5.1: Συγκριτική επισκόπηση GNN μοντέλων για συστάσεις

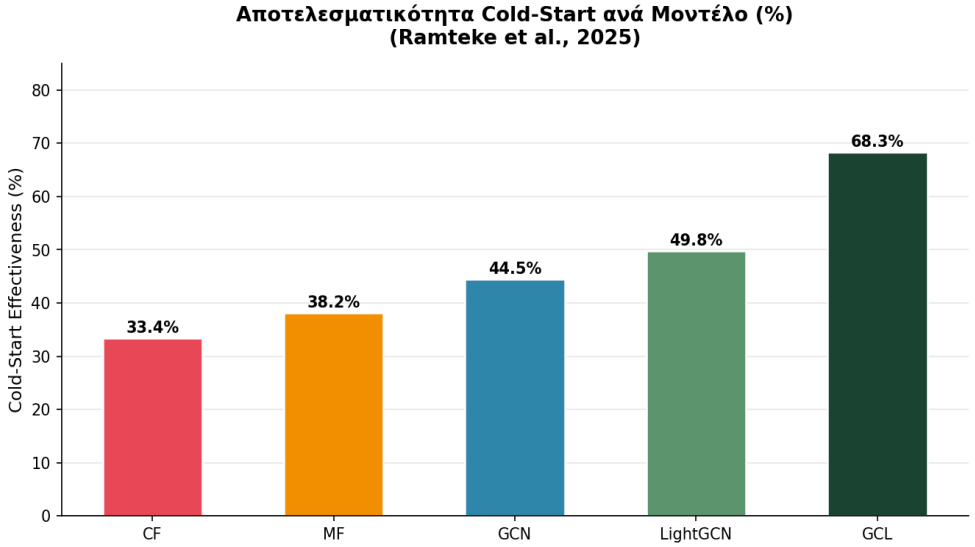
Μοντέλο	Τύπος Γράφου	Βασική Καινοτομία	Κύριο Πλεονέκτημα
NGCF	Δίγραφος χρήστη-αντικειμένου	Message passing με embeddings	Πρώτο GCN για RS
LightGCN	Δίγραφος χρήστη-αντικειμένου	Αφαίρεση μη-γραμμικότητας & μετασχηματισμών	Απλότητα, scalability
GAT	Γενικοί γράφοι	Attention βάρη σε γείτονες	Επιλεκτική συνάθροιση
SR-GNN	Session γράφος αντικειμένων	Gated GNN σε session sequences	Session-based σύσταση
KGCN	Γνωσιακός γράφος	Διάδοση σε knowledge graph	Σημασιολογική πλαισίωση αντικειμένων
HDHP	Ετερογενής υπεργράφος	HGCN για υψηλής τάξης σχέσεις	Πλούσιες κοινωνικές αλληλεπιδράσεις
MycGNN	Δυναμικός δίγραφος	Mycelium-inspired εξερεύνηση	Ποικιλομορφία συστάσεων
GCL / SGL	Δίγραφος + augmentation	Contrastive learning σε γράφο	Αντιμετώπιση αραιότητας & bias

Το Διάγραμμα 5.1 συγκρίνει το Hit Rate@10 παραδοσιακών μεθόδων, GCN-based μοντέλων και Graph Contrastive Learning. Το GCL επιτυγχάνει HR@10 = 82.4%, ξεπερνώντας κατά 6.2 μονάδες το LightGCN (76.2%) και κατά 21 μονάδες το CF (61.4%). Η διαφορά μεγαλώνει περαιτέρω σε σενάρια long-tail αντικειμένων και cold-start χρηστών, όπου τα παραδοσιακά μοντέλα υστερούν δραματικά.



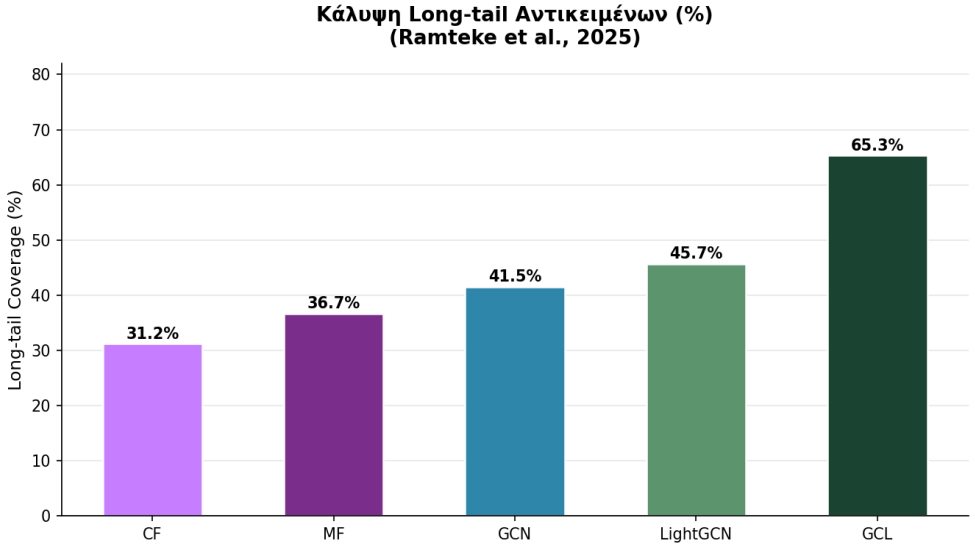
Διάγραμμα 5.1: Hit Rate@10 σύγκριση CF, GNN-based και GCL

Το Διάγραμμα 5.2 αναδεικνύει την υπεροχή του GCL στην αντιμετώπιση cold-start. Ενώ το CF αποτυγχάνει κατά 66.6% σε cold-start σενάρια, το GCL φτάνει στο 68.3% αποτελεσματικότητα (σχεδόν διπλάσιο αποτέλεσμα) χάρη στην αυτο-επιβλεπόμενη εκπαίδευση που δεν εξαρτάται από επαρκές ιστορικό. Η κλίση της καμπύλης επιτάχυνσης αυξάνεται ιδιαίτερα από GCN (44.5%) σε LightGCN (49.8%) και GCL (68.3%), υποδηλώνοντας ότι η απλοποίηση της γραφικής διάδοσης συμβάλλει θετικά.



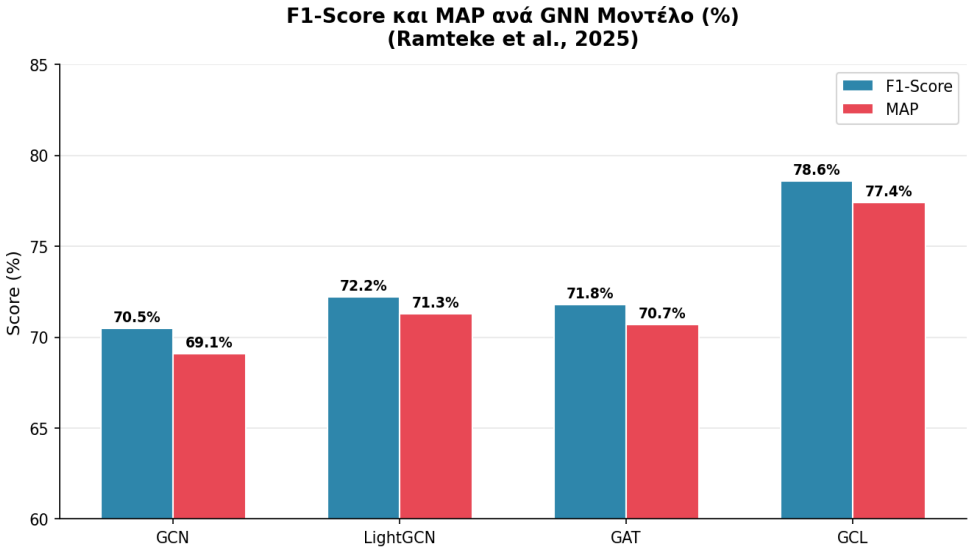
Διάγραμμα 5.2: Αποτελεσματικότητα cold-start ανά μοντέλο

Το Διάγραμμα 5.3 παρουσιάζει την κάλυψη long-tail αντικειμένων: δηλαδή αντικειμένων με ελάχιστες αλληλεπιδράσεις που τα παραδοσιακά μοντέλα αδυνατούν να συστήσουν. Το GCL επιτυγχάνει 65.3%, ενώ το CF μόλις 31.2%. Αυτή η βελτίωση σημαίνει ότι το GCL μπορεί να αξιοποιήσει ουσιαστικά τη μακριά ουρά αντικειμένων, μειώνοντας τη μεροληψία δημοτικότητας (popularity bias) και αυξάνοντας τη diversity των συστάσεων.



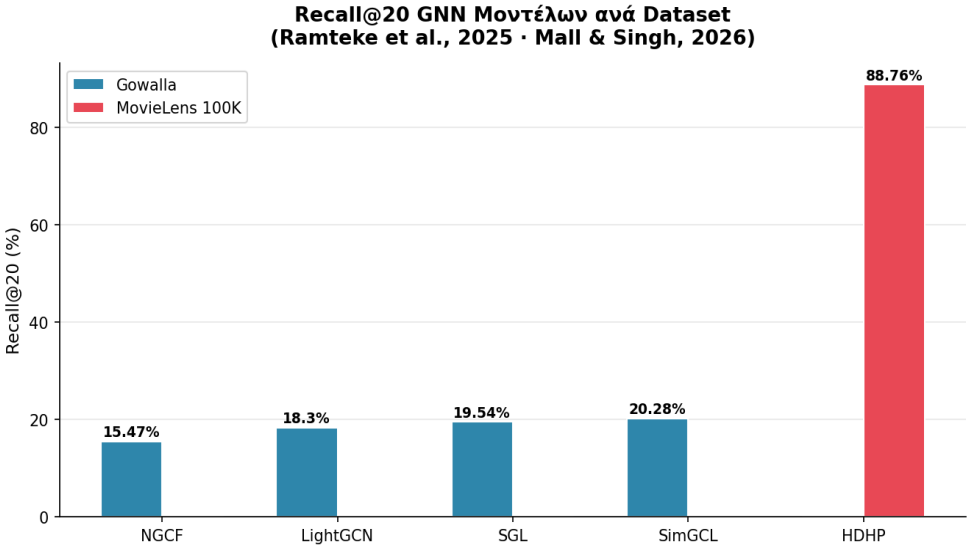
Διάγραμμα 5.3: Κάλυψη long-tail αντικειμένων ανά μοντέλο

Το Διάγραμμα 5.4 συγκρίνει F1-Score και MAP για τέσσερα GNN μοντέλα. Παρατηρείται ότι GCN και GAT έχουν παρόμοια απόδοση (F1: 70.5% vs 71.8%), γεγονός που υποδηλώνει ότι ο μηχανισμός attention δεν αρκεί από μόνος του για ουσιαστική βελτίωση. Η απλοποίηση του LightGCN δίνει μόλις οριακή βελτίωση (72.2%), ενώ το GCL που προσθέτει αντιπαραθετική εκπαίδευση εκτοξεύει την απόδοση στο 78.6% F1: αύξηση 8.1 μονάδων έναντι GCN.



Διάγραμμα 5.4: Σύγκριση F1-Score και MAP μεταξύ GNN μοντέλων

Το Διάγραμμα 5.5 συγκρίνει Recall@20 κυριότερων GNN μοντέλων σε Gowalla και MovieLens 100K. Στο Gowalla, το SimGCL (20.28%) ξεπερνά όλα τα baseline μοντέλα, με 31% βελτίωση έναντι NGCF (15.47%). Στο MovieLens 100K, το HDHP Hypergraph επιτυγχάνει εξαιρετικό Recall@20 = 88.76%, αναδεικνύοντας τη δύναμη των ετερογενών υπεργράφων για κοινωνική σύσταση.



Διάγραμμα 5.5: Recall@20 GNN μοντέλων σε Gowalla και MovieLens 100K

5.9 Cross-domain σύσταση με GNN

Η Cross-domain Recommendation (CDR) αποτελεί μια σημαντική επέκταση των GNN-based συστημάτων που αντιμετωπίζει τον περιορισμό των μεθόδων ενός τομέα: την αδυναμία εκμετάλλευσης γνώσης από σχετικούς τομείς. Η βασική ιδέα είναι ότι οι προτιμήσεις χρήστη σε έναν τομέα (π.χ. ανάγνωση βιβλίων) μπορούν να βελτιώσουν τις συστάσεις σε έναν άλλο σχετικό τομέα (π.χ. ταινίες ή μουσική), αξιοποιώντας κοινές λανθάνουσες αναπαραστάσεις ή μηχανισμούς μεταφοράς γνώσης.

Ο Patil & Basavaraju (2026) πρότειναν ένα υβριδικό πλαίσιο GNN για cross-domain σύσταση που συνδυάζει δύο συνιστώσες. Η πρώτη χρησιμοποιεί κοινά (shared) και domain-specific embeddings με global bias terms, εκπαιδευόμενα σε δύο στάδια: supervised warm-up με MSE loss και στη συνέχεια BPR fine-tuning για βελτιστοποίηση κατάταξης. Η δεύτερη

βασίζεται σε αρχιτεκτονική εμπνευσμένη από το BiTGCF, που χρησιμοποιεί ετερογενή bipartite γράφο χρήστη-αντικειμένου για εμφάνιση εξαρτήσεων υψηλής τάξης και domain alignment.

Πειράματα στα datasets Amazon και Yelp κατέδειξαν ότι το μοντέλο επιτυγχάνει Recall@10 = 0.0989 στο Amazon και 0.0141 στο Yelp (αποτελέσματα σταθερά ανώτερα από single-domain LightGCN και cross-domain baselines όπως DeepCDR και CDCL. Ο Πίνακας 5.2 παρουσιάζει συνοπτικά τα αποτελέσματα. Σημαντική παρατήρηση είναι ότι η υπεροχή του μοντέλου είναι πιο έντονη στο dataset Yelp, όπου η data sparsity είναι μεγαλύτερη) ακριβώς εκεί όπου η μεταφορά γνώσης από το Amazon είναι πιο απαραίτητη.

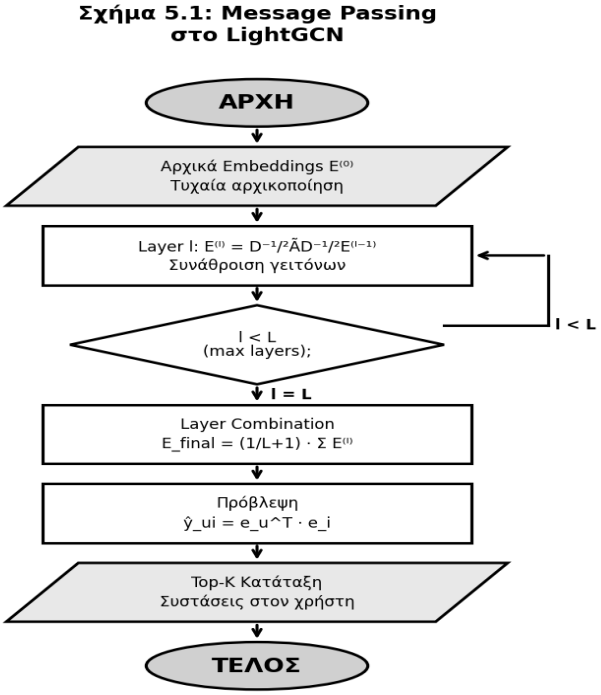
Πίνακας 5.2: Αποτελέσματα cross-domain μοντέλων σε Amazon και Yelp

Μοντέλο	Dataset	Recall@10	NDCG@10	Προσέγγιση
MF (baseline)	Amazon	0.0412	0.0198	Ενιαίος τομέας
LightGCN (baseline)	Amazon	0.0698	0.0321	Ενιαίος τομέας
DeepCDR	Amazon	0.0821	0.0389	Cross-domain DNN
CDCL	Amazon	0.0901	0.0418	Contrastive CDR
BiTGCF-Hybrid	Amazon	0.0989	0.0451	GNN + Embedding CDR
MF (baseline)	Yelp	0.0072	0.0031	Ενιαίος τομέας
CDCL	Yelp	0.0112	0.0051	Contrastive CDR
BiTGCF-Hybrid	Yelp	0.0141	0.0064	GNN + Embedding CDR

Οι κυριότερες προκλήσεις της CDR αφορούν τρεις περιοχές. Πρώτον, το πρόβλημα domain shift: οι κατανομές αντικειμένων και χρηστών διαφέρουν μεταξύ τομέων, καθιστώντας δύσκολη τη μεταφορά αναπαραστάσεων χωρίς υποβάθμιση. Δεύτερον, η ιδιωτικότητα δεδομένων: σε πολλές πραγματικές εφαρμογές, τα δεδομένα διαφορετικών τομέων ανήκουν σε διαφορετικές εταιρείες και δεν μπορούν να κοινοποιηθούν. Αυτό ωθεί στην ανάπτυξη federated cross-domain learning μεθόδων. Τρίτον, η scalability: η ταυτόχρονη εκπαίδευση σε πολλαπλούς τομείς αυξάνει σημαντικά τον υπολογιστικό φόρτο.

5.10 Flowchart διαδικασίας message passing LightGCN

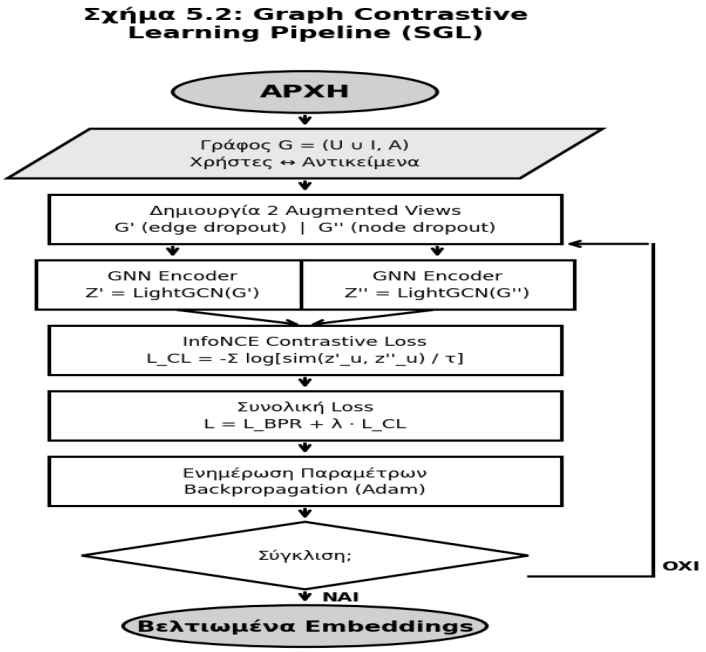
Το Σχήμα 5.1 απεικονίζει τη διαδικασία message passing στο LightGCN. Ξεκινώντας από τα αρχικά embeddings $E^{(0)}$, κάθε επίπεδο εκτελεί κανονικοποιημένη συνάθροιση γειτόνων χωρίς μη γραμμικές ενεργοποιήσεις. Μετά από L επίπεδα, τα embeddings κάθε επιπέδου συνδυάζονται με mean pooling στο τελικό embedding E_final. Η πρόβλεψη υπολογίζεται ως εσωτερικό γινόμενο των τελικών embeddings χρήστη και αντικειμένου.



Σχήμα 5.1: Διαδικασία message passing στο LightGCN

5.11 Pipeline Graph Contrastive Learning

Το Σχήμα 5.2 παρουσιάζει το pipeline του Graph Contrastive Learning (SGL). Από τον αρχικό γράφο G δημιουργούνται δύο augmented views G' και G'' μέσω ανεξάρτητων τεχνικών επαύξησης (edge dropout, node dropout). Κάθε view περνά από ξεχωριστό GNN encoder, παράγοντας embeddings Z' και Z''. Η InfoNCE contrastive loss ωθεί τα embeddings του ίδιου κόμβου να είναι κοντά και διαφορετικών κόμβων να απομακρύνονται. Η συνολική loss συνδυάζει BPR και contrastive term.



Σχήμα 5.2: Pipeline Graph Contrastive Learning: δημιουργία views, encoding, contrastive loss

5.12 Το πρόβλημα του Over-smoothing στα GNN

Μια από τις πιο γνωστές προκλήσεις στην εκπαίδευση βαθιών GNN αρχιτεκτονικών είναι το φαινόμενο του over-smoothing (υπερ-εξομάλυνσης). Καθώς αυξάνεται ο αριθμός των επιπέδων διάδοσης μηνυμάτων (message passing layers), οι αναπαραστάσεις (embeddings) διαφορετικών κόμβων τείνουν να συγκλίνουν προς όμοιες τιμές, χάνοντας τη διακριτική τους ικανότητα. Στο πλαίσιο των συστημάτων συστάσεων, αυτό σημαίνει ότι μετά από αρκετά επίπεδα διάδοσης, οι αναπαραστάσεις χρηστών και αντικειμένων καθίστανται δυσδιάκριτες μεταξύ τους, υποβαθμίζοντας σημαντικά την ποιότητα των προτεινόμενων συστάσεων.

Το φαινόμενο συνδέεται άμεσα με τις ιδιότητες κανονικοποίησης (regularization) του LightGCN (Παράρτημα Γ, Ενότητα Γ.2): καθώς ο κανονικοποιημένος πίνακας γειτνίασης $\tilde{A}$ εφαρμόζεται επανειλημμένα, οι αναπαραστάσεις συγκλίνουν προς τον κυρίαρχο ιδιοχώρο του $\tilde{A}$, ο οποίος αντιστοιχεί σε χαμηλής συχνότητας («ομαλά») σήματα στον γράφο. Στην πράξη, μοντέλα με περισσότερα από 3-4 επίπεδα συνήθως εμφανίζουν σημαντική υποβάθμιση απόδοσης, γεγονός που εξηγεί εν μέρει γιατί το LightGCN διατηρεί σχετικά μικρό αριθμό επιπέδων.

Έχουν προταθεί διάφορες τεχνικές αντιμετώπισης: οι συνδέσεις παράλειψης (residual/skip connections), που επιτρέπουν στο τελικό embedding να συνδυάζει πληροφορία από όλα τα επίπεδα αντί μόνο από το τελευταίο: προσέγγιση που υιοθετεί το ίδιο το LightGCN μέσω της σταθμισμένης άθροισης επιπέδων· η regularization PairNorm, που διατηρεί σταθερή τη συνολική διασπορά (variance) των αναπαραστάσεων ανά επίπεδο· και η τεχνική DropEdge, που αφαιρεί τυχαία ακμές του γράφου σε κάθε εποχή εκπαίδευσης, λειτουργώντας ως μορφή κανονικοποίησης που περιορίζει την υπερβολική διάδοση πληροφορίας.

5.13 Ετερογενή GNN για συστάσεις

Τα μοντέλα GNN που παρουσιάστηκαν στις προηγούμενες ενότητες, όπως το LightGCN (Ενότητα 5.2), υποθέτουν έναν ομογενή γράφο με έναν τύπο κόμβου

(χρήστες/αντικείμενα) και έναν τύπο ακμής (αλληλεπίδραση). Στην πράξη, ωστόσο, πολλές εφαρμογές σύστασης διαθέτουν πλουσιότερη δομή γράφου, με πολλαπλούς τύπους κόμβων (π.χ. χρήστες, αντικείμενα, κατηγορίες, ετικέτες, δημιουργοί περιεχομένου) και πολλαπλούς τύπους σχέσεων μεταξύ τους. Τα ετερογενή GNN (Heterogeneous GNNs) επεκτείνουν τον μηχανισμό διάδοσης μηνυμάτων ώστε να χειρίζεται διαφορετικούς τύπους κόμβων και ακμών με ξεχωριστούς πίνακες μετασχηματισμού ανά τύπο σχέσης.

Η προσέγγιση αυτή επιτρέπει την ενσωμάτωση πλουσιότερης πλευρικής πληροφορίας (side information) απευθείας στη δομή του γράφου, αντί να προστίθεται ως πρόσθετο χαρακτηριστικό εισόδου, όπως συμβαίνει σε απλούστερες υβριδικές προσεγγίσεις. Στο πεδίο της φαρμακευτικής φροντίδας που αναλύθηκε στην Ενότητα 7.9, για παράδειγμα, τα ετερογενή δίκτυα φαρμάκων-ασθενειών-πρωτεϊνών αξιοποιούν ακριβώς αυτή την ιδιότητα για τον εντοπισμό έμμεσων σχέσεων μέσω πολλαπλών τύπων κόμβων.

5.14 Κλιμακωσιμότητα: τεχνικές δειγματοληψίας γειτονιάς

Η πλήρης διάδοση μηνυμάτων σε γράφους με δισεκατομμύρια ακμές, όπως αυτοί που εμφανίζονται σε πλατφόρμες κλίμακας Pinterest ή μεγάλων δικτύων ηλεκτρονικού εμπορίου, είναι υπολογιστικά απαγορευτική. Το GraphSAGE εισήγαγε την τεχνική της δειγματοληψίας γειτονιάς σταθερού μεγέθους (fixed-size neighbor sampling), όπου σε κάθε επίπεδο διάδοσης λαμβάνεται υπόψη μόνο ένα τυχαίο υποσύνολο σταθερού μεγέθους των γειτόνων κάθε κόμβου, αντί για το σύνολο των γειτόνων, μειώνοντας δραστικά την υπολογιστική πολυπλοκότητα από εκθετική σε γραμμική ως προς τον αριθμό των επιπέδων.

Το PinSage, η βιομηχανική εφαρμογή αυτής της ιδέας στο Pinterest, συνδυάζει τη δειγματοληψία γειτονιάς με τυχαίους περιπάτους (random walks) με βάρη βασισμένα στην προσωποποιημένη κατάταξη PageRank, ώστε να δίνεται προτεραιότητα σε γείτονες με μεγαλύτερη δομική σημασία. Τέτοιες τεχνικές δειγματοληψίας αποτελούν προαπαιτούμενο για την πρακτική εφαρμογή των GNN αρχιτεκτονικών που παρουσιάστηκαν στο παρόν κεφάλαιο σε περιβάλλοντα παραγωγής πραγματικής κλίμακας, όπως συζητήθηκε και στην Ενότητα 9.7 για τις αρχιτεκτονικές δύο σταδίων.

5.15 Διαχυτικά μοντέλα (Diffusion Models) σε γράφους σύστασης

Μια πρόσφατη και ταχύτατα αναπτυσσόμενη κατεύθυνση έρευνας αφορά την εφαρμογή διαχυτικών μοντέλων (diffusion models), που έχουν αποδειχθεί ιδιαίτερα επιτυχημένα στη γεννητική μοντελοποίηση εικόνας, στο πλαίσιο των συστημάτων συστάσεων βασισμένων σε γράφο. Η βασική ιδέα είναι η προσθήκη ελεγχόμενου θορύβου στις αναπαραστάσεις χρηστών ή αντικειμένων μέσω μιας διαδικασίας πρόσθιας διάχυσης (forward diffusion), και η εκπαίδευση ενός νευρωνικού δικτύου να αντιστρέφει αυτή τη διαδικασία (reverse diffusion), μαθαίνοντας έτσι να ανακατασκευάζει καθαρές, υψηλής ποιότητας αναπαραστάσεις από θορυβώδεις εκδοχές τους.

Στο πλαίσιο της σύστασης, η προσέγγιση αυτή έχει εφαρμοστεί σε συνδυασμό με γνωσιακή ενίσχυση (knowledge-enhanced) και ιεραρχική αξιοποίηση πληροφορίας γράφου, με στόχο την παραγωγή πιο ανθεκτικών αναπαραστάσεων απέναντι σε αραιά ή θορυβώδη δεδομένα αλληλεπίδρασης, σε μια λογική που θυμίζει τη στοχαστική κωδικοποίηση των VAE (Ενότητα 4.13), αλλά με διαφορετική μαθηματική θεμελίωση βασισμένη σε στοχαστικές διαφορικές εξισώσεις. Η υπολογιστική απαίτηση των διαχυτικών μοντέλων παραμένει σημαντικά υψηλότερη από τις συμβατικές αρχιτεκτονικές GNN, θέτοντας ζητήματα πρακτικής εφαρμοσιμότητας ανάλογα με αυτά που συζητήθηκαν στην Ενότητα 8.13 περί ενεργειακής αποδοτικότητας.

5.16 Ασαφής ενίσχυση γράφων (Fuzzy-enhanced GCN)

Μια πρόσθετη παραλλαγή των συνελικτικών γραφικών δικτύων ενσωματώνει αρχές ασαφούς λογικής (fuzzy logic) στον υπολογισμό των βαρών των ακμών του γράφου, αντί να βασίζεται αποκλειστικά σε δυαδική ή σταθερή regularization όπως στο βασικό LightGCN (Ενότητα 5.2). Σε ένα ασαφές GCN με μεταβλητά βάρη (fuzzy-enhanced variable-weight GCN), η ισχύς της σχέσης μεταξύ δύο κόμβων εκφράζεται μέσω συναρτήσεων συμμετοχής ασαφούς λογικής, οι οποίες αποτυπώνουν με μεγαλύτερη ευελιξία τον βαθμό αβεβαιότητας ή μερικής συσχέτισης που χαρακτηρίζει πολλές πραγματικές σχέσεις χρήστη-αντικειμένου.

Αυτή η προσέγγιση είναι ιδιαίτερα χρήσιμη σε σενάρια όπου η ίδια η αλληλεπίδραση δεν είναι δυαδική (όπως αγόρασε/δεν αγόρασε) αλλά εμπεριέχει διαβαθμίσεις έντασης ή

αβεβαιότητας: για παράδειγμα, η μερική ολοκλήρωση ενός βίντεο ή η σύντομη παραμονή σε μια σελίδα προϊόντος. Η ενσωμάτωση ασαφούς λογικής προσφέρει έναν εναλλακτικό μηχανισμό κωδικοποίησης αυτής της αβεβαιότητας σε σχέση με την πιθανοτική προσέγγιση των Variational Autoencoders που παρουσιάστηκε στην Ενότητα 4.13, με τα δύο πλαίσια να μοιράζονται τον κοινό στόχο της ρητής αναπαράστασης αβεβαιότητας στο μοντέλο σύστασης.

5.17 Δυναμικά γραφικά νευρωνικά δίκτυα (Dynamic GNNs)

Τα μοντέλα GNN που παρουσιάστηκαν στις προηγούμενες ενότητες υποθέτουν έναν στατικό γράφο, ο οποίος ανανεώνεται περιοδικά μέσω πλήρους επανεκπαίδευσης. Τα δυναμικά GNN (dynamic GNNs), αντίθετα, μοντελοποιούν ρητά την εξέλιξη του γράφου αλληλεπιδράσεων στο χρόνο, ενσωματώνοντας νέες ακμές (αλληλεπιδράσεις) σταδιακά και ενημερώνοντας τις αναπαραστάσεις των κόμβων με αυξητικό τρόπο, αντί να απαιτούν πλήρη επανυπολογισμό σε κάθε νέα αλληλεπίδραση.

Αυτή η προσέγγιση συνδυάζει στοιχεία από τη σειριακή σύσταση (Ενότητα 4.6) με τη δομική πληροφορία του γράφου, επιτρέποντας στο μοντέλο να αποτυπώνει ταυτόχρονα τόσο τη χρονική εξέλιξη των προτιμήσεων ενός μεμονωμένου χρήστη όσο και τις δομικές σχέσεις μεταξύ χρηστών και αντικειμένων στο επίπεδο ολόκληρου του γράφου. Η υπολογιστική πολυπλοκότητα της αυξητικής ενημέρωσης παραμένει σημαντικά μικρότερη από την πλήρη επανεκπαίδευση, καθιστώντας τα δυναμικά GNN ιδιαίτερα ελκυστική επιλογή για πλατφόρμες παραγωγής με υψηλό ρυθμό άφιξης νέων δεδομένων, σε αντιστοιχία με τις ανησυχίες ενεργειακής αποδοτικότητας που συζητήθηκαν στην Ενότητα 8.13.

5.18 Σύνοψη

Πιο εντυπωσιακή από την απλή απόδοση των GNN βρίσκω την αλλαγή οπτικής που εισάγουν: το NGCF και το LightGCN δεν προσθέτουν απλώς μια ακόμη αρχιτεκτονική, αλλά αναδιατυπώνουν το ίδιο το πρόβλημα της σύστασης ως πρόβλημα πάνω σε γράφο, μέσω message passing σε δίγραφους χρήστη-αντικειμένου. Τα GAT προσθέτουν attention στη διαδικασία αυτή, τα GNN σε Knowledge Graphs την εμπλουτίζουν σημασιολογικά, οι ετερογενείς γράφοι και υπεργράφοι επιτρέπουν σύλληψη πολύπλοκων σχέσεων, και το Graph Contrastive Learning αντιμετωπίζει αραιότητα και μεροληψία. Ωστόσο το over-smoothing, ο θόρυβος στη διάδοση και η δυσκολία σύλληψης μακρών εξαρτήσεων παραμένουν, κατά την εκτίμησή μου, πραγματικά όρια της προσέγγισης -- και όχι απλώς «λεπτομέρειες υλοποίησης» -- κάτι που εξηγεί γιατί η έρευνα στράφηκε προς υβριδικές αρχιτεκτονικές GNN-Transformer.

6. Ειδικές κατηγορίες συστημάτων συστάσεων

Πέραν των γενικών μεθόδων που αναλύθηκαν στα προηγούμενα κεφάλαια, η ερευνητική κοινότητα έχει αναπτύξει εξειδικευμένες κατηγορίες συστημάτων συστάσεων που αντιμετωπίζουν συγκεκριμένες προκλήσεις ή εκμεταλλεύονται πρόσθετες πηγές πληροφορίας. Στις επόμενες ενότητες εξετάζονται έξι από τις σημαντικότερες εξειδικευμένες κατηγορίες: τα σειριακά συστήματα (sequential recommendation), τα συστήματα πολλαπλής συμπεριφοράς (multi-behavior), τα συστήματα ευαίσθητα στο πλαίσιο (context-aware), τα συστήματα εμπιστοσύνης (trust-based), τα ερμηνεύσιμα συστήματα (explainable) και τα συστήματα διαλόγου (conversational). Κοινός παρονομαστής τους είναι η προσπάθεια να κατανοηθεί ο χρήστης πληρέστερα, πηγαίνοντας πέρα από τα απλά ιστορικά δεδομένα αξιολογήσεων.

6.1 Σειριακή σύσταση (Sequential recommendation)

Η σειριακή σύσταση (sequential recommendation) ξεκινά από μια θεμελιώδη παρατήρηση: οι προτιμήσεις ενός χρήστη δεν είναι στατικές, αλλά εξελίσσονται δυναμικά στο χρόνο, και η σειρά με την οποία ο χρήστης έχει αλληλεπιδράσει με αντικείμενα στο παρελθόν περιέχει σημαντική πληροφορία για τις μελλοντικές του προτιμήσεις. Η ακολουθία των αλληλεπιδράσεων αντιμετωπίζεται ως μια «πρόταση» σε αναλογία με τις λέξεις σε ένα κείμενο, και οι τεχνικές επεξεργασίας ακολουθιών από το πεδίο της επεξεργασίας φυσικής γλώσσας εφαρμόζονται για την πρόβλεψη του επόμενου αντικειμένου ενδιαφέροντος.

Ο Chen et al. (2026) ορίζουν ότι στο πρόβλημα της σειριακής σύστασης δίνεται μια χρονικά ταξινομημένη ακολουθία αλληλεπιδράσεων ενός χρήστη $S_u = [i_1, i_2, ..., i_t]$ και ο στόχος είναι η πρόβλεψη του πιο πιθανού επόμενου αντικειμένου i_{t+1}. Αυτό διαφέρει από τα

γενικά συστήματα σύστασης, όπου ο στόχος είναι η παραγωγή μιας λίστας συστάσεων χωρίς να ληφθεί υπόψη η χρονική σειρά των αλληλεπιδράσεων. Η σειριακή φύση επιτρέπει την ανάδειξη βραχυπρόθεσμων ενδιαφερόντων του χρήστη, συμπληρώνοντας τις μακροπρόθεσμες προτιμήσεις που συλλαμβάνουν οι παραδοσιακές μέθοδοι.

6.1.1 Μοντέλα και προκλήσεις

Τα κυριότερα μοντέλα σειριακής σύστασης αναλύθηκαν στο Κεφάλαιο 4 (RNN-based: GRU4Rec, HRNN) και στο Κεφάλαιο 5 (GNN-based: SR-GNN, SURGE). Στην κατηγορία των Transformer-based μοντέλων ξεχωρίζουν το SASRec και το BERT4Rec. Το SASRec χρησιμοποιεί μονόδρομο self-attention για να μοντελοποιεί σειριακά πρότυπα, ενώ το BERT4Rec χρησιμοποιεί αμφίδρομο self-attention με masked item prediction, επιτυγχάνοντας βαθύτερη κατανόηση του ιστορικού χρήστη. Μια σχετικά νέα κατηγορία είναι τα session-based μοντέλα, όπου ο χρήστης δεν αναγνωρίζεται ρητά αλλά οι συστάσεις γίνονται βάσει της τρέχουσας σύντομης συνόδου (session).

Κύρια πρόκληση των σειριακών συστημάτων είναι η σύλληψη τόσο βραχυπρόθεσμων όσο και μακροπρόθεσμων εξαρτήσεων μέσα σε μια ακολουθία. Τα απλά RNN πάσχουν από το πρόβλημα της εξαφανιζόμενης βαθμίδας για μακρές ακολουθίες, ενώ τα Transformer χειρίζονται μακρά ιστορικά αποτελεσματικά αλλά με υψηλότερο υπολογιστικό κόστος. Η ενσωμάτωση χρονικών σημάτων (π.χ. χρονοσφραγίδων (timestamps)) αποτελεί ενεργό ερευνητική κατεύθυνση.

6.2 Σύσταση πολλαπλής συμπεριφοράς (Multi-behavior recommendation)

Στις περισσότερες πραγματικές εφαρμογές, οι χρήστες αλληλεπιδρούν με αντικείμενα με διαφορετικούς τρόπους που εκφράζουν διαφορετικά επίπεδα ενδιαφέροντος. Σε μια πλατφόρμα ηλεκτρονικού εμπορίου, για παράδειγμα, ένας χρήστης μπορεί να εξετάσει ένα προϊόν (view), να το προσθέσει στο καλάθι (add-to-cart), να το κάνει αγαπημένο (favourite) ή να το αγοράσει (purchase). Κάθε τύπος αλληλεπίδρασης αντικατοπτρίζει διαφορετική πρόθεση και η αξιοποίηση αυτής της ετερογένειας είναι ο στόχος των συστημάτων σύστασης πολλαπλής συμπεριφοράς.

Οι Chen et al. (2026) ορίζουν το πρόβλημα της multi-behavior sequential recommendation (MBSR) ως τη μοντελοποίηση τόσο της σειριακής φύσης των αλληλεπιδράσεων όσο και της ετερογένειάς τους. Τα κύρια ερευνητικά ερωτήματα είναι: πώς να μοντελοποιηθούν οι σχέσεις μεταξύ διαφορετικών τύπων συμπεριφοράς, πώς να συνδυαστούν βραχυπρόθεσμες και μακροπρόθεσμες προτιμήσεις που πηγάζουν από διαφορετικές συμπεριφορές, και πώς να αντιμετωπιστεί ο θόρυβος που εισάγουν ορισμένοι τύποι αλληλεπίδρασης. Η προσέγγιση αυτή είναι ιδιαίτερα αποτελεσματική γιατί η συμπεριφορά-στόχος (π.χ. αγορά) έχει πολύ λιγότερες παρατηρήσεις σε σχέση με βοηθητικές συμπεριφορές (π.χ. εξέταση), με αποτέλεσμα η μόνο χρήση της συμπεριφοράς-στόχου να οδηγεί σε ακραία data sparsity.

Μεταξύ των σύγχρονων μοντέλων MBSR, τα GNN-based αποδεικνύονται ιδιαίτερα αποτελεσματικά: το MBGCN (Multi-Behavior Graph Convolutional Network) μοντελοποιεί κάθε τύπο συμπεριφοράς ως ξεχωριστό γράφο και μαθαίνει αναπαραστάσεις μέσω μηνυμάτων στο σύνολο των γράφων. Τα Transformer-based μοντέλα χρησιμοποιούν ξεχωριστούς μηχανισμούς attention για κάθε τύπο συμπεριφοράς, ενώ εξειδικευμένοι μηχανισμοί cross-behavior attention επιτρέπουν τη μεταφορά γνώσης μεταξύ τύπων. Πειράματα σε πραγματικά datasets (Taobao, Beibei) επιβεβαιώνουν ότι τα MBSR μοντέλα ξεπερνούν σταθερά τα αντίστοιχα single-behavior μοντέλα.

6.3 Συστήματα ευαίσθητα στο πλαίσιο (Context-aware recommendation)

Τα Συστήματα Συστάσεων Ευαίσθητα στο Πλαίσιο (Context-Aware Recommender Systems, CARS) αναγνωρίζουν ότι η προτίμηση ενός χρήστη για ένα αντικείμενο δεν εξαρτάται μόνο από τα χαρακτηριστικά του χρήστη και του αντικειμένου, αλλά και από το πλαίσιο (context) της αλληλεπίδρασης. Ο Chafiki et al. (2026) ορίζουν το πλαίσιο ως «οποιαδήποτε πληροφορία που μπορεί να χρησιμοποιηθεί για να χαρακτηρίσει την κατάσταση μιας οντότητας», συμπεριλαμβάνοντας χωρικές πληροφορίες (τοποθεσία), χρονικές πληροφορίες (ώρα, ημέρα, εποχή), κοινωνικές πληροφορίες (παρουσία άλλων ατόμων), συναισθηματικές πληροφορίες (διάθεση) και περιβαλλοντικές πληροφορίες (καιρός, θόρυβος).

6.3.1 Στρατηγικές ενσωμάτωσης πλαισίου

Η βιβλιογραφία αναγνωρίζει τρεις κύριες στρατηγικές για την ενσωμάτωση πλαισιακής πληροφορίας: την προ-φιλτράρισμα (pre-filtering), όπου το πλαίσιο χρησιμοποιείται για τη φιλτράρισμα των διαθέσιμων δεδομένων πριν την εφαρμογή του αλγορίθμου σύστασης, το μετα-φιλτράρισμα (post-filtering), όπου τα αποτελέσματα μιας συμβατικής σύστασης αναπροσαρμόζονται βάσει πλαισίου, και την πλαισιακή μοντελοποίηση (contextual modeling), όπου το πλαίσιο ενσωματώνεται άμεσα στον αλγόριθμο σύστασης. Η τρίτη προσέγγιση, αν και πιο σύνθετη, τείνει να δίνει τα καλύτερα αποτελέσματα σε πειράματα.

Ένα από τα πιο πρόσφατα και αποτελεσματικά CARS μοντέλα είναι το MCCRS (Multi-type Context-aware Conversational Recommender System) των Zou et al. (2026), το οποίο συνδυάζει δομημένες πληροφορίες (γνωσιακοί γράφοι), αδόμητες πληροφορίες ιστορικού διαλόγου και κριτικές αντικειμένων μέσω μιας αρχιτεκτονικής mixture-of-experts. Κάθε εμπειρογνώμονας (expert) εξειδικεύεται σε έναν τύπο πλαισιακής πληροφορίας, ενώ ένα ChairBot συντονίζει τα αποτελέσματα. Τα πειράματα έδειξαν σημαντικά υψηλότερη απόδοση σε σχέση με τα υφιστάμενα baseline μοντέλα. Ο Chafiki et al. (2026) επισημαίνουν ότι παρά την ποικιλία των πλαισιακών τύπων και μεθοδολογιών, οι επίμονες προκλήσεις παραμένουν: η διαχείριση ετερογενούς πλαισίου, η scalability και η αξιολόγηση.

6.4 Συστήματα βασισμένα στην εμπιστοσύνη (Trust-based recommendation)

Τα συστήματα σύστασης βασισμένα στην εμπιστοσύνη (trust-based recommender systems) αξιοποιούν διαπροσωπικές σχέσεις εμπιστοσύνης μεταξύ χρηστών για τη βελτίωση της ποιότητας των συστάσεων. Η θεμελιώδης παρατήρηση είναι ότι οι χρήστες τείνουν να αποδέχονται συστάσεις από ανθρώπους που εμπιστεύονται περισσότερο από ό,τι από αγνώστους, ακόμα και αν οι τελευταίοι έχουν παρόμοιες προτιμήσεις. Οι Akdim et al. (2026) ορίζουν ότι «η εμπιστοσύνη, σε αυτό το πλαίσιο, αναφέρεται στην πιθανότητα ένας χρήστης να υιοθετήσει τις προτιμήσεις άλλων χρηστών του δικτύου που θεωρεί αξιόπιστους».

Η ενσωμάτωση εμπιστοσύνης στα συστήματα σύστασης αντιμετωπίζει αρκετές τεχνικές προκλήσεις. Πρώτον, η εμπιστοσύνη δεν είναι πάντα ρητά δηλωμένη: συχνά πρέπει να συναχθεί από έμμεσα δεδομένα, όπως αξιολογήσεις, σχόλια ή κοινές συνδέσεις σε κοινωνικά δίκτυα. Δεύτερον, η εμπιστοσύνη είναι ασύμμετρη: ο χρήστης Α μπορεί να εμπιστεύεται τον Β χωρίς το αντίστροφο. Τρίτον, η εμπιστοσύνη μεταξύ χρηστών χωρίς άμεση σχέση πρέπει να υπολογίζεται μέσω διάδοσης στο δίκτυο εμπιστοσύνης (trust propagation). Τεχνικές όπως το TrustWalker και το TidalTrust υπολογίζουν εμπιστοσύνη μεταξύ μη άμεσα συνδεδεμένων χρηστών μέσω διαδρομών στο δίκτυο.

Τα σύγχρονα trust-based μοντέλα αξιοποιούν GNN για τη μοντελοποίηση δικτύων εμπιστοσύνης. Το DiffNet μοντελοποιεί τη διάδοση κοινωνικής επιρροής ως διαδικασία διάχυσης (diffusion) στο κοινωνικό γράφο, ενώ το SEPT (Social Enhanced Pre-Training) εκπαιδεύει αναπαραστάσεις χρηστών αξιοποιώντας ταυτόχρονα δεδομένα αλληλεπίδρασης και κοινωνικές σχέσεις. Εμπειρικές μελέτες επιβεβαιώνουν ότι οι trust-based προσεγγίσεις παρέχουν ακριβέστερες συστάσεις σε σχέση με τις αντίστοιχες χωρίς εμπιστοσύνη, ιδιαίτερα σε σενάρια αραιών δεδομένων.

6.5 Ερμηνεύσιμα συστήματα συστάσεων (Explainable recommendation)

Τα ερμηνεύσιμα συστήματα συστάσεων (explainable recommender systems) στοχεύουν όχι μόνο στην παραγωγή ακριβών συστάσεων, αλλά και στην παροχή ερμηνειών που εξηγούν στον χρήστη γιατί προτείνεται ένα συγκεκριμένο αντικείμενο. Αυτή η ικανότητα έχει τεράστια σημασία για την αποδοχή των συστάσεων από τους χρήστες, τη δημιουργία εμπιστοσύνης στο σύστημα και τη συμμόρφωση με κανονισμούς διαφάνειας, ειδικά σε κρίσιμους τομείς όπως η υγεία, η χρηματοοικονομική και η εκπαίδευση.

Η βιβλιογραφία αναγνωρίζει διαφορετικές μορφές ερμηνειών: κειμενικές εξηγήσεις σε φυσική γλώσσα, εξηγήσεις βάσει χαρακτηριστικών (feature-based), εξηγήσεις βάσει παραδειγμάτων (example-based) και οπτικοποιήσεις. Οι Zarzour et al. (2022) παρουσιάζουν ένα σύστημα που χρησιμοποιεί βαθιά νευρωνικά δίκτυα για την πρόβλεψη θετικών κριτικών στα πλαίσια ερμηνεύσιμης σύστασης, επιτυγχάνοντας καλύτερα αποτελέσματα σε precision, recall και F1 σε σχέση με τα baseline μοντέλα σε dataset Amazon.

6.5.1 Μεγάλα Γλωσσικά Μοντέλα στην ερμηνεύσιμη σύσταση

Η εμφάνιση των Μεγάλων Γλωσσικών Μοντέλων (Large Language Models, LLMs) άνοιξε νέους δρόμους για την ερμηνεύσιμη σύσταση. Το STLLM-Rec (Self-Training LLM Recommendation) των Li et al. (2026) αξιοποιεί τις δυνατότητες συλλογισμού των LLMs για να παράγει ρητές αιτιολογήσεις πίσω από κάθε σύσταση, αντιμετωπίζοντας τρεις κρίσιμες προκλήσεις: την τάση του μοντέλου να «ταιριάζει» κριτικές αντί να παράγει πραγματικές αιτιολογήσεις, τους περιορισμούς γενίκευσης σε zero-shot σενάρια και την έλλειψη δεδομένων εκπαίδευσης υψηλής ποιότητας για εξηγήσεις. Η μέθοδος self-training ενσωματώνει λεπτομερή σήματα ανταμοιβής (reward signals) που καθοδηγούν τη διαδικασία συλλογισμού. Τα πειράματα έδειξαν ότι το STLLM-Rec επιτυγχάνει υψηλή ακρίβεια σύστασης και ισχυρή ερμηνευσιμότητα ταυτόχρονα.

6.6 Συστήματα με επίγνωση δικαιοσύνης (Fairness-aware recommendation)

Καθώς τα συστήματα συστάσεων γίνονται ολοένα πιο διαδεδομένα, αναδύεται ένα κρίσιμο ζήτημα: η μεροληψία (bias). Τα μοντέλα εκπαιδευμένα σε δεδομένα πραγματικού κόσμου κληρονομούν και ενισχύουν τις κοινωνικές ανισότητες που υπάρχουν στα δεδομένα, με αποτέλεσμα να εμφανίζουν άνισες επιδόσεις για διαφορετικές ομάδες χρηστών ή να ευνοούν συστηματικά ορισμένες κατηγορίες αντικειμένων. Το ζήτημα της δικαιοσύνης (fairness) αφορά τόσο την πλευρά των χρηστών (user-side fairness) όσο και την πλευρά των παρόχων (provider-side fairness).

Οι Yang et al. (2026) αναλύουν λεπτομερώς τους δύο τύπους μεροληψίας που ενισχύει ο μηχανισμός μεταβίβασης μηνυμάτων των GNN: την attribute bias (μεροληψία ως προς χαρακτηριστικά κόμβων, π.χ. φύλο ή ηλικία) και την topology bias (μεροληψία που πηγάζει από την ανισότητα συνδέσεων στο γράφο). Το προτεινόμενο μοντέλο FairDA (Fair Dual-path Alignment) αντιμετωπίζει και τους δύο τύπους ταυτόχρονα: το Fairness-Oriented Distillation (FOD) παράγει αναπαραστάσεις χρηστών χαμηλής μεροληψίας μέσω distillation από ένα δάσκαλο-μοντέλο εκπαιδευμένο σε αποχαρακτηρισμένα δεδομένα, ενώ το Difference-Aware Consistency Regularization (DACR) ρυθμίζει δυναμικά τις αναπαραστάσεις αντικειμένων διαφορετικών ομάδων. Πειράματα σε δύο πραγματικά datasets επιβεβαιώνουν ότι το FairDA επιτυγχάνει ανώτερη ισορροπία μεταξύ ακρίβειας σύστασης και δικαιοσύνης.

6.7 Συστήματα σύστασης μέσω διαλόγου (Conversational recommendation)

Τα Συστήματα Συστάσεων Μέσω Διαλόγου (Conversational Recommender Systems, CRS) αντιπροσωπεύουν μια ριζικά διαφορετική προσέγγιση στη διαδικασία σύστασης: αντί για μια μονόδρομη παραγωγή λίστας, το σύστημα εμπλέκεται σε διαδραστική συνομιλία με τον χρήστη σε φυσική γλώσσα, αποσαφηνίζοντας σταδιακά τις προτιμήσεις και τις ανάγκες του. Αυτό επιτρέπει μια πολύ πιο εκλεπτυσμένη κατανόηση του χρήστη, ιδιαίτερα σε σενάρια cold start όπου ελάχιστο ιστορικό αλληλεπίδρασης είναι διαθέσιμο.

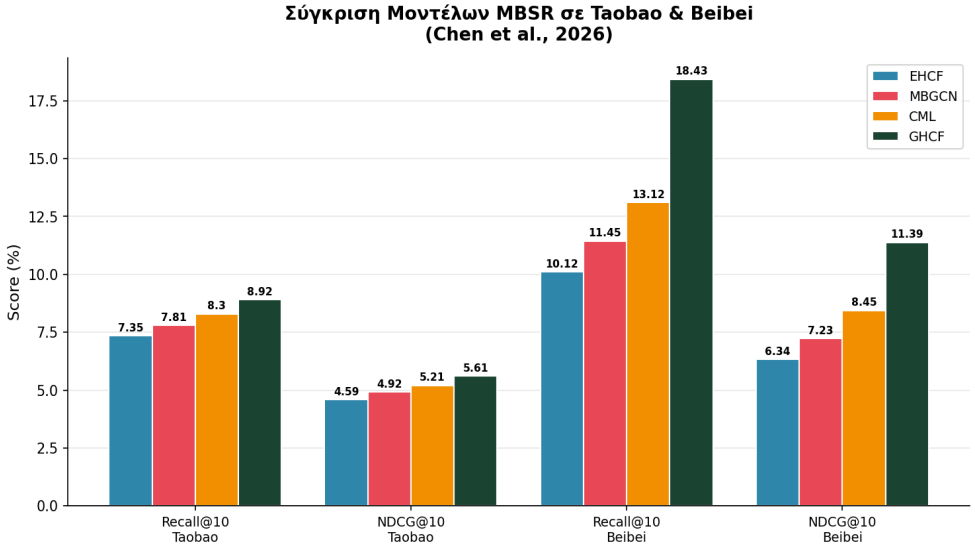
Το MCCRS των Zou et al. (2026) αποτελεί χαρακτηριστικό παράδειγμα σύγχρονου CRS. Αντιμετωπίζει την ετερογένεια των πλαισιακών πηγών πληροφορίας (γνωσιακοί γράφοι (δομημένα), ιστορικό συνομιλίας και κριτικές (αδόμητα)) χρησιμοποιώντας μια αρχιτεκτονική mixture-of-experts. Κάθε expert εξειδικεύεται σε έναν τύπο πλαισιακής πληροφορίας και ένα ChairBot συντονίζει τα αποτελέσματα για την τελική σύσταση. Αυτή η αρχιτεκτονική επιτρέπει επίσης ιχνηλασιμότητα (traceability): είναι εύκολο να αναγνωριστεί ποιος expert ήταν υπεύθυνος για ένα αποτέλεσμα. Πειράματα επιβεβαιώνουν σημαντικά υψηλότερη απόδοση έναντι των baseline μοντέλων.

Παρά την πρόοδο, τα CRS αντιμετωπίζουν σημαντικές ανοιχτές προκλήσεις: η κατανόηση σύντομων και ασαφών ερωτημάτων χρήστη, η διαχείριση πολλαπλών στόχων συνομιλίας (πληροφόρηση vs. σύσταση vs. εξήγηση), και η αποφυγή επαναλαμβανόμενων ή ανεπαρκών ερωτήσεων προς τον χρήστη αποτελούν ενεργά ερευνητικά ερωτήματα. Τα LLM-based CRS αρχίζουν να αντιμετωπίζουν αυτές τις προκλήσεις αξιοποιώντας την πλούσια πρότερη γνώση των γλωσσικών μοντέλων.

Πίνακας 6.1: Συνοπτική σύγκριση ειδικών κατηγοριών συστημάτων συστάσεων

Κατηγορία	Βασική Ιδέα	Κύρια Πρόκληση	Ενδεικτικά Μοντέλα
Sequential	Σύσταση βάσει χρονικής ακολουθίας αλληλεπιδράσεων	Μακροπρόθεσμες εξαρτήσεις	GRU4Rec, SASRec, BERT4Rec
Multi-behavior	Αξιοποίηση ετερογενών τύπων συμπεριφοράς	Μοντελοποίηση σχέσεων μεταξύ συμπεριφορών	MBGCN, CML, GHCF
Context-aware	Ενσωμάτωση πλαισιακών παραγόντων στη σύσταση	Διαχείριση ετερογενούς πλαισίου	CARS2, DeepCoNN, MCCRS
Trust-based	Αξιοποίηση σχέσεων εμπιστοσύνης μεταξύ χρηστών	Διάδοση εμπιστοσύνης σε μεγάλα δίκτυα	TrustMF, SocialMF, DiffNet
Explainable	Παροχή ερμηνευσιμών αιτιολογήσεων σύστασης	Ισορροπία ακρίβειας – ερμηνευσιμότητας	Att2Seq, PETER, STLLM-Rec
Fairness-aware	Δίκαιη μεταχείριση ομάδων χρηστών / αντικειμένων	Αφαίρεση μεροληψίας χωρίς απώλεια ακρίβειας	FairDA, FairGNN, FairRec
Conversational	Διαδραστική σύσταση μέσω φυσικής γλώσσας	Κατανόηση σύντομων διαλόγων	MCCRS, EAR, CRSLab

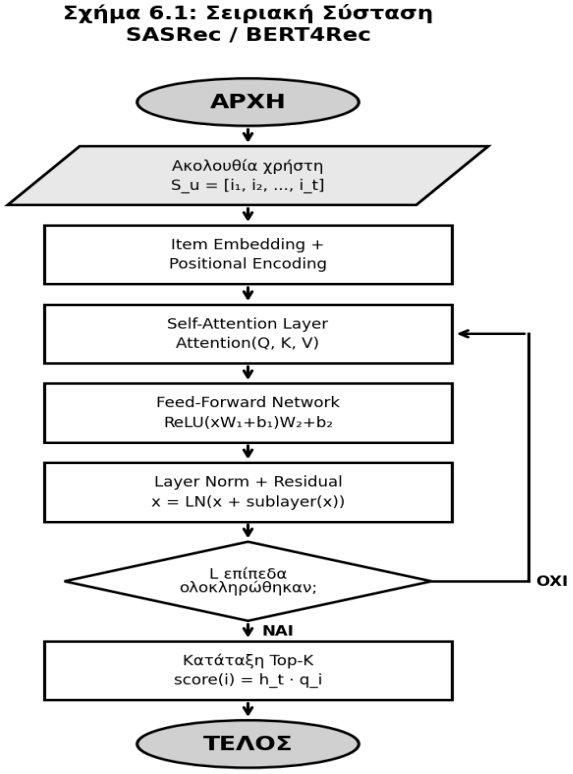
Το Διάγραμμα 6.1 συγκρίνει τέσσερα μοντέλα Multi-Behavior Sequential Recommendation σε Taobao και Beibei. Το GHCF (Hypergraph) επιτυγχάνει σταθερά τα καλύτερα αποτελέσματα: Recall@10 = 8.92% στο Taobao και 18.43% στο Beibei. Η μεγάλη διαφορά απόδοσης μεταξύ Taobao και Beibei οφείλεται στην υψηλότερη πυκνότητα δεδομένων του Beibei (1.93% vs 0.14%), που επιτρέπει στα μοντέλα να αξιοποιήσουν πλήρως την πολυπλοκότητα των πολλαπλών συμπεριφορών.



Διάγραμμα 6.1: Σύγκριση μοντέλων MBSR σε Taobao και Beibei

6.8 Flowchart σειριακής σύστασης SASRec

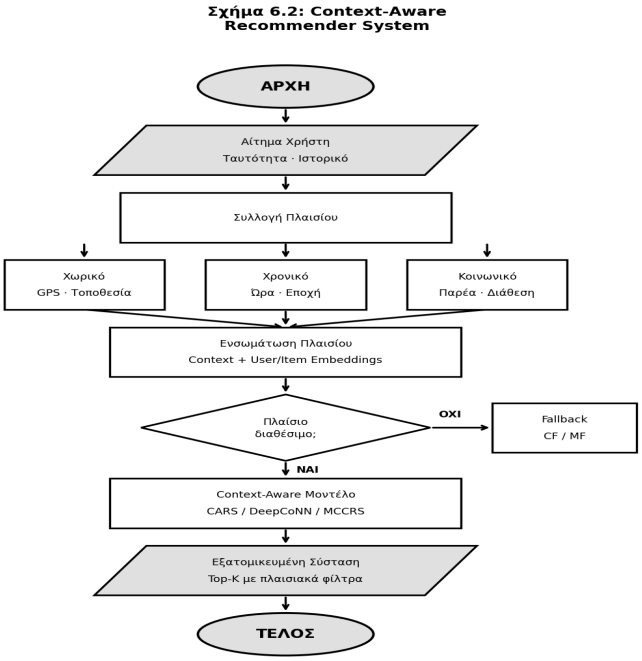
Το Σχήμα 6.1 περιγράφει τη διαδικασία σειριακής σύστασης βάσει Transformer (SASRec/BERT4Rec). Η ακολουθία αλληλεπιδράσεων S_u = [i_1, ..., i_t] κωδικοποιείται με item embeddings και positional encoding. Ακολουθούν L επίπεδα self-attention και Feed-Forward Networks με residual connections και layer normalization. Η διαδικασία επαναλαμβάνεται για κάθε επίπεδο, και το τελικό hidden state χρησιμοποιείται για κατάταξη και παραγωγή Top-K συστάσεων.



Σχήμα 6.1: Flowchart σειριακής σύστασης βάσει Transformer (SASRec/BERT4Rec)

6.9 Ροή αποφάσεων context-aware RS

Το Σχήμα 6.2 απεικονίζει τη ροή αποφάσεων ενός Context-Aware Recommender System. Κατά την υποβολή αιτήματος σύστασης, το σύστημα συλλέγει παράλληλα τρεις κατηγορίες πλαισιακής πληροφορίας: χωρική (τοποθεσία GPS), χρονική (ώρα, ημέρα, εποχή) και κοινωνική (παρέα, διάθεση). Εάν το πλαίσιο είναι διαθέσιμο, ενσωματώνεται στο μοντέλο σύστασης· αλλιώς το σύστημα υποβαθμίζεται σε κλασικό CF/MF.



Σχήμα 6.2: Ροή αποφάσεων Context-Aware Recommender System

6.10 Σύσταση πολλαπλών ενδιαφερόμενων μερών (Multi-stakeholder recommendation)

Οι περισσότερες κλασικές προσεγγίσεις σύστασης βελτιστοποιούνται αποκλειστικά ως προς την ικανοποίηση του τελικού χρήστη (consumer-centric). Στην πράξη, ωστόσο, οι πραγματικές πλατφόρμες σύστασης εμπλέκουν πολλαπλά ενδιαφερόμενα μέρη με εν μέρει ανταγωνιστικούς στόχους: τους καταναλωτές, που αναζητούν σχετικό περιεχόμενο· τους παρόχους περιεχομένου ή προϊόντων (providers), που επιδιώκουν δίκαιη προβολή των αντικειμένων τους· και την ίδια την πλατφόρμα, που στοχεύει σε μακροπρόθεσμη βιωσιμότητα του οικοσυστήματος αγοράς.

Η σύσταση πολλαπλών ενδιαφερόμενων μερών (multi-stakeholder recommendation) διατυπώνει το πρόβλημα ως πολυκριτηριακή βελτιστοποίηση, αναζητώντας σύστασεις που ικανοποιούν ταυτόχρονα, ή τουλάχιστον ισορροπούν, τους στόχους όλων των εμπλεκόμενων μερών. Αυτό σχετίζεται στενά με τη δικαιοσύνη παρόχων (provider fairness) που αναλύθηκε στην Ενότητα 6.6, αλλά επεκτείνεται σε ευρύτερα ζητήματα, όπως η ισόρροπη προβολή νέων έναντι καθιερωμένων παρόχων και η αποφυγή μεροληψιών αναπαράστασης (representation bias) που μπορούν να προκύψουν από μονομερή βελτιστοποίηση μόνο ως προς τη συμπεριφορά κατανάλωσης.

Τεχνικά, η προσέγγιση αυτή συνήθως υλοποιείται μέσω επανακατάταξης (re-ranking) μετά την αρχική παραγωγή λίστας υποψήφιων αντικειμένων, εφαρμόζοντας περιορισμούς ή ποινές στη συνάρτηση βαθμολόγησης ώστε να ενσωματωθούν στόχοι πέραν της ακρίβειας πρόβλεψης. Η ισορροπία μεταξύ ακρίβειας και πολυμερούς δικαιοσύνης παραμένει, κατά τη γνώμη μου, το πιο δύσκολο ανοιχτό ζήτημα της κατηγορίας, ιδιαίτερα σε δυναμικά περιβάλλοντα όπου οι στόχοι των ενδιαφερόμενων μερών μεταβάλλονται με την πάροδο του χρόνου.

6.11 Τεχνική υλοποίηση διαλογικών συστημάτων σύστασης (NLU και πολυγυρικές αλληλεπιδράσεις)

Πέραν της γενικής επισκόπησης της Ενότητας 6.7, η τεχνική υλοποίηση των CRS βασίζεται σε μια αλληλεπιδραστική διαδικασία πολλαπλών γύρων (multi-turn) μεταξύ χρήστη και συστήματος, στην οποία το σύστημα μπορεί να θέσει διευκρινιστικές ερωτήσεις για να περιορίσει τον χώρο αναζήτησης πριν διατυπώσει μια σύσταση, αντί να βασίζεται αποκλειστικά σε παθητική ανάλυση ιστορικών δεδομένων. Η προσέγγιση αυτή είναι ιδιαίτερα χρήσιμη σε σενάρια όπου οι προτιμήσεις του χρήστη είναι ασαφείς, εξαρτώνται από το άμεσο πλαίσιο (π.χ. τη διάθεση της στιγμής), ή όταν ο χώρος αντικειμένων είναι ιδιαίτερα μεγάλος και πολυδιάστατος.

Τεχνικά, ένα τυπικό σύστημα διαλόγου-σύστασης συνδυάζει έναν μηχανισμό κατανόησης φυσικής γλώσσας (NLU) για την εξαγωγή προτιμήσεων από τις απαντήσεις του χρήστη, μια πολιτική διαλόγου (dialogue policy) που αποφασίζει πότε να ρωτήσει διευκρινιστικά και πότε να προτείνει, και έναν πυρήνα σύστασης που ενσωματώνει την συσσωρευμένη πληροφορία του διαλόγου, συχνά υλοποιημένο με αρχιτεκτονικές Transformer παρόμοιες με αυτές της Ενότητας 4.5. Η ενσωμάτωση πλαισιακής πληροφορίας (context-aware τεχνικές, Κεφάλαιο 6.3) ενισχύει περαιτέρω την ποιότητα αυτών των συστημάτων.

6.12 Διαφοροποίηση σειριακών και session-based συστάσεων

Παρότι οι όροι «σειριακή σύσταση» (sequential recommendation) και «σύσταση βάσει συνόδου» (session-based recommendation) χρησιμοποιούνται συχνά εναλλακτικά στη βιβλιογραφία, υπάρχει μια σημαντική εννοιολογική διαφορά. Τα σειριακά μοντέλα, όπως το SASRec (Ενότητα 4.5), υποθέτουν γνωστή και σταθερή ταυτότητα χρήστη στο πέρασμα του χρόνου, μοντελοποιώντας τη μακροπρόθεσμη εξέλιξη των προτιμήσεών του μέσα σε πολλαπλές, χρονικά διακριτές συνόδους αλληλεπίδρασης.

Τα μοντέλα βάσει συνόδου (session-based), αντίθετα, αντιμετωπίζουν κάθε σύνοδο αλληλεπίδρασης ως ανεξάρτητη μονάδα ανάλυσης, συχνά χωρίς να γνωρίζουν ή να βασίζονται στην ταυτότητα του χρήστη: ένα σενάριο ιδιαίτερα συχνό σε ανώνυμη περιήγηση ηλεκτρονικού εμπορίου. Αρχιτεκτονικές GNN-based, όπως το SR-GNN που αναφέρθηκε στην Ενότητα 6.1, μοντελοποιούν τη σύνοδο ως γράφο μεταβάσεων μεταξύ αντικειμένων, αποτυπώνοντας πολύπλοκες, μη γραμμικές μεταβατικές σχέσεις που τα απλά σειριακά μοντέλα ενδέχεται να παραβλέψουν.

6.13 Εξατομικευμένη σύσταση με μοντελοποίηση χαρακτηριστικών προσωπικότητας

Μια λιγότερο διερευνημένη, αλλά αναπτυσσόμενη κατηγορία συστημάτων σύστασης ενσωματώνει ρητά χαρακτηριστικά προσωπικότητας του χρήστη (π.χ. βάσει του μοντέλου Big Five) ως πρόσθετη πηγή πληροφορίας, πέρα από το ιστορικό αλληλεπιδράσεων. Η υπόθεση εργασίας είναι ότι χρήστες με όμοια χαρακτηριστικά προσωπικότητας τείνουν να εμφανίζουν παρόμοια μοτίβα προτίμησης ακόμη και σε διαφορετικά είδη περιεχομένου, παρέχοντας έτσι ένα συμπληρωματικό σήμα ιδιαίτερα χρήσιμο σε σενάρια cold start (Ενότητα 2.11), όπου το ιστορικό αλληλεπιδράσεων είναι ανεπαρκές.

Τεχνικά, τα χαρακτηριστικά προσωπικότητας ενσωματώνονται είτε ως πρόσθετα χαρακτηριστικά εισόδου σε υβριδικά μοντέλα τύπου LSTM-SVM, είτε ως ξεχωριστός λανθάνων χώρος που συντήκεται με τις κλασικές αναπαραστάσεις χρήστη-αντικειμένου μέσω μηχανισμών προσοχής, παρόμοια με τη λογική σύντηξης πολλαπλών κλάδων που περιγράφηκε για το NeuMF στην Ενότητα 4.1. Η προσέγγιση αυτή εγείρει ωστόσο πρόσθετα ζητήματα ιδιωτικότητας, καθώς η συναγωγή χαρακτηριστικών προσωπικότητας από συμπεριφορικά δεδομένα θεωρείται ιδιαίτερα ευαίσθητη επεξεργασία δεδομένων προσωπικού χαρακτήρα, συνδέοντας το θέμα άμεσα με τη συζήτηση περί ιδιωτικότητας της Ενότητας 8.3.

6.14 Σύσταση με ενίσχυση συναισθηματικής ανάλυσης

Μια συμπληρωματική κατηγορία ειδικών συστημάτων σύστασης ενσωματώνει συναισθηματική ανάλυση (sentiment analysis) κειμενικών κριτικών ή σχολίων χρηστών ως πρόσθετη πηγή σήματος προτίμησης, πέρα από τις αριθμητικές βαθμολογίες ή τις δυαδικές ενδείξεις αλληλεπίδρασης. Η υπόθεση εργασίας είναι ότι το συναισθηματικό περιεχόμενο μιας γραπτής κριτικής (π.χ. ενθουσιασμός, απογοήτευση) μεταφέρει λεπτότερες αποχρώσεις προτίμησης από μια απλή αριθμητική βαθμολογία, ιδιαίτερα σε περιπτώσεις όπου η βαθμολογία

δεν αντικατοπτρίζει πλήρως το κείμενο της κριτικής.

Τεχνικά, η συναισθηματική βαθμολογία εξάγεται συνήθως μέσω προ-εκπαιδευμένων μοντέλων επεξεργασίας φυσικής γλώσσας και ενσωματώνεται είτε ως πρόσθετο χαρακτηριστικό εισόδου σε ένα μοντέλο παραγοντοποίησης πίνακα, είτε ως ξεχωριστός όρος στη συνάρτηση κόστους που εξαναγκάζει τις εκμαθημένες αναπαραστάσεις να είναι συνεπείς με το συναισθηματικό περιεχόμενο των αντίστοιχων κριτικών. Ο συνδυασμός αυτός με τεχνικές προσαρμοζόμενης μάθησης σε χρονικές σειρές (time-series adaptive learning) επιτρέπει επιπλέον την παρακολούθηση της εξέλιξης της συναισθηματικής στάσης ενός χρήστη απέναντι σε μια κατηγορία προϊόντων με την πάροδο του χρόνου.

6.15 Σύσταση σε πολυτροπικά (multi-modal) δεδομένα

Πολλές σύγχρονες εφαρμογές σύστασης διαθέτουν αντικείμενα που περιγράφονται από πολλαπλούς τύπους δεδομένων ταυτόχρονα (κείμενο, εικόνα, ήχο ή βίντεο) μια ιδιότητα που εκμεταλλεύονται τα πολυτροπικά (multi-modal) συστήματα σύστασης. Στον τομέα της μόδας, για παράδειγμα, η οπτική εμφάνιση ενός ρούχου (Πίνακας 7.3) αποτελεί εξίσου σημαντικό σήμα προτίμησης όσο και το ιστορικό αγορών, ενώ στη μουσική, τα ακουστικά χαρακτηριστικά συμπληρώνουν τα σήματα σειριακής συμπεριφοράς ακρόασης.

Η ενσωμάτωση πολυτροπικών δεδομένων απαιτεί ξεχωριστούς κωδικοποιητές (encoders) ανά τύπο δεδομένων (π.χ. συνελικτικά δίκτυα για εικόνα, Transformer για κείμενο) με τις προκύπτουσες αναπαραστάσεις να συντήκονται σε κοινό latent space μέσω μηχανισμών προσοχής, παρόμοιων με αυτούς που περιγράφηκαν για το NeuMF (Ενότητα 4.1) και τα ετερογενή GNN (Ενότητα 5.13). Η πρόκληση της ευθυγράμμισης (alignment) μεταξύ τροπικοτήτων με διαφορετική στατιστική φύση παραμένει τεχνικά ανεπίλυτη, κυρίως επειδή δεν υπάρχει ακόμη κοινώς αποδεκτή μετρική για να αξιολογηθεί πόσο καλά «ταιριάζουν» δύο τροπικότητες, ιδίως όταν δεν είναι πάντα διαθέσιμες για κάθε αντικείμενο του καταλόγου.

6.16 Σύσταση με περιορισμούς ποσόστωσης και επιχειρηματικούς κανόνες

Σε πραγματικά συστήματα παραγωγής, οι καθαρά αλγοριθμικές συστάσεις συνδυάζονται συχνά με ρητούς επιχειρηματικούς κανόνες και περιορισμούς ποσόστωσης (business rules and quota constraints), οι οποίοι αντικατοπτρίζουν εμπορικούς ή στρατηγικούς στόχους που δεν προκύπτουν φυσικά από τη βελτιστοποίηση ακρίβειας. Για παράδειγμα, μια πλατφόρμα ενδέχεται να απαιτεί ότι τουλάχιστον ένα ποσοστό των προτεινόμενων αντικειμένων σε κάθε λίστα προέρχεται από προωθούμενο ή νέο περιεχόμενο, ανεξάρτητα από το αν αυτό μεγιστοποιεί τη στατιστική πιθανότητα αλληλεπίδρασης.

Η ενσωμάτωση τέτοιων περιορισμών υλοποιείται τυπικά στο στάδιο της επανακατάταξης (re-ranking), αξιοποιώντας τεχνικές βελτιστοποίησης με περιορισμούς (constrained optimization) πάνω στις αρχικές βαθμολογίες που παράγει το μοντέλο σύστασης, σε μια λογική ανάλογη με αυτή των τεχνικών πολλαπλών ενδιαφερόμενων μερών της Ενότητας 6.10. Η ισορροπία μεταξύ αυστηρής τήρησης επιχειρηματικών κανόνων και διατήρησης ικανοποιητικής ποιότητας σύστασης από την οπτική του χρήστη αποτελεί συνεχή πρακτική πρόκληση για τις ομάδες μηχανικής μάθησης που λειτουργούν τέτοια συστήματα σε περιβάλλον παραγωγής.

6.17 Σύνοψη

Οι έξι αυτές κατηγορίες δεν είναι, κατά τη γνώμη μου, εξίσου ώριμες: τα σειριακά και τα context-aware συστήματα έχουν ήδη εδραιωθεί σε παραγωγικά περιβάλλοντα, ενώ τα conversational και τα fairness-aware παραμένουν πιο πειραματικά. Αντιμετωπίζουν συγκεκριμένες πτυχές του προβλήματος πέρα από τις κλασικές και deep learning μεθόδους. Τα σειριακά συστήματα εκμεταλλεύονται τη χρονική δομή των αλληλεπιδράσεων, τα multi-behavior συστήματα αξιοποιούν την ετερογένεια των τύπων συμπεριφοράς, τα context-aware ενσωματώνουν πλαισιακή πληροφορία, τα trust-based χρησιμοποιούν κοινωνικές σχέσεις εμπιστοσύνης, τα explainable παρέχουν διαφάνεια στη διαδικασία σύστασης, τα fairness-aware αντιμετωπίζουν ζητήματα μεροληψίας και τα conversational επιτρέπουν διαδραστική εκλέπτυνση προτιμήσεων. Η κατανόηση αυτών των κατηγοριών είναι απαραίτητη για την ανάπτυξη συστημάτων που ανταποκρίνονται στις πραγματικές ανάγκες χρηστών σε σύνθετα και ετερογενή περιβάλλοντα.

7. Τομείς εφαρμογής συστημάτων συστάσεων

Τα συστήματα συστάσεων δεν αποτελούν απλώς ακαδημαϊκό αντικείμενο έρευνας: είναι ήδη διάχυτα σε καθημερινές ψηφιακές εφαρμογές, διαμορφώνοντας σε μεγάλο βαθμό τον τρόπο με τον οποίο οι χρήστες ανακαλύπτουν προϊόντα, προορισμούς, εκπαιδευτικό υλικό, ψυχαγωγία και πληροφόρηση. Στις επόμενες ενότητες εξετάζονται οι σημαντικότεροι τομείς εφαρμογής: ηλεκτρονικό εμπόριο, τουρισμό και ταξίδια, εκπαίδευση, τρόφιμα και διατροφή, κοινωνικά δίκτυα και ψυχαγωγία. Σε κάθε τομέα αναλύονται οι ιδιαίτερες προκλήσεις, οι επικρατούσες τεχνικές και τα εμπειρικά αποτελέσματα από τη βιβλιογραφία.

7.1 Ηλεκτρονικό εμπόριο (E-commerce) και κοινωνικό ηλεκτρονικό εμπόριο

Το ηλεκτρονικό εμπόριο αποτελεί τον πιο ώριμο και εμπορικά κρίσιμο τομέα εφαρμογής των συστημάτων συστάσεων. Ο παγκόσμιος όγκος αγορών ηλεκτρονικού εμπορίου ανήλθε περίπου σε 5,14 τρισεκατομμύρια δολάρια το 2024 και εκτιμάται ότι θα ξεπεράσει τα 8 τρισεκατομμύρια έως το 2027, με τα συστήματα συστάσεων να διαδραματίζουν κεντρικό ρόλο στη διασύνδεση χρηστών με προϊόντα. Πλατφόρμες όπως η Amazon, η Taobao και η Pinduoduo έχουν ενσωματώσει εξελιγμένα συστήματα συστάσεων που αυξάνουν τον μέσο χρόνο παραμονής κατά 37% και οδηγούν σε μετρήσιμη αύξηση μετατροπών και εσόδων.

Ο Yu & Aviles (2025) πρότειναν ένα μοντέλο για πλατφόρμες ηλεκτρονικού εμπορίου που συνδυάζει Convolutional Neural Network (CNN) για εξαγωγή χαρακτηριστικών με Long Short-Term Memory (LSTM) για μοντελοποίηση χρονοσειράς, ενισχυμένο με μηχανισμό attention. Το σύστημα εφαρμόστηκε σε πραγματικό dataset ηλεκτρονικού εμπορίου και παρουσίασε ακρίβεια 89%, αισθητά υψηλότερη από το collaborative filtering (72%) και τη matrix factorization (78%). Σε εμπορικούς δείκτες, το σύστημα αύξησε σταδιακά το conversion rate στο 24% την 30ή ημέρα, ανέβασε τις μέσες ημερήσιες πωλήσεις κατά 25% και βελτίωσε τον δείκτη διατήρησης χρηστών κατά 15 ποσοστιαίες μονάδες. Ο Huang (2022) εφάρμοσε CNN-based σύστημα συστάσεων στο dataset Alibaba, καταδεικνύοντας την υπεροχή του έναντι αλγορίθμων συνεργατικής διήθησης και κανόνων συσχέτισης σε ρεαλιστικά σενάρια κλίμακας.

Μια ιδιαίτερα ενδιαφέρουσα υποκατηγορία είναι το κοινωνικό ηλεκτρονικό εμπόριο (social e-commerce), όπου οι κοινωνικές σχέσεις χρηστών αξιοποιούνται για βελτίωση των συστάσεων. Ο Zhai (2025) κατασκεύασε έναν ετερογενή GNN γράφο που μοντελοποιεί χρήστες, προϊόντα και τα χαρακτηριστικά τους ως κόμβους διαφορετικών τύπων, εφαρμόζοντας graph convolution για εκμάθηση πλούσιων αναπαραστάσεων. Τα αποτελέσματα ανέδειξαν σημαντική βελτίωση ακρίβειας και ανάκλησης σε σχέση με τα baseline μοντέλα CF, επιβεβαιώνοντας ότι η μοντελοποίηση των πολύπλοκων σχέσεων μεταξύ χρηστών, προϊόντων και κοινωνικών δεσμών οδηγεί σε ποιοτικά ανώτερες συστάσεις σε περιβάλλοντα ηλεκτρονικού εμπορίου.

Στο πλαίσιο του κοινωνικού ηλεκτρονικού εμπορίου (social e-commerce), όπου οι αγοραστικές αποφάσεις επηρεάζονται έντονα από κοινωνικές συστάσεις φίλων και επιρροητών, τα γραφικά νευρωνικά δίκτυα έχουν εφαρμοστεί για τη βελτιστοποίηση αλγορίθμων σύστασης που λαμβάνουν υπόψη ταυτόχρονα το γράφο κοινωνικών σχέσεων και το γράφο αλληλεπιδράσεων προϊόντων. Η ενσωμάτωση αυτών των δύο διακριτών πηγών πληροφορίας σε μια ενιαία αρχιτεκτονική γράφου ακολουθεί τη λογική των ετερογενών GNN που παρουσιάστηκε στην Ενότητα 5.13.

Παράλληλα, στα ευρύτερα πλατφόρμες ηλεκτρονικού εμπορίου, συστήματα σύστασης βασισμένα σε βαθιά μάθηση έχουν σχεδιαστεί ειδικά για την υποστήριξη της λήψης απόφασης αγοράς σε πραγματικό χρόνο, ενσωματώνοντας σήματα όπως η διάρκεια παραμονής σε σελίδα προϊόντος και η αλληλουχία περιήγησης εντός της ίδιας συνόδου, σε αντιστοιχία με τις τεχνικές session-based σύστασης που αναλύθηκαν στην Ενότητα 6.12.

7.2 Τουρισμός και ταξίδια

Ο τουρισμός αποτελεί ένα ιδιαίτερα σύνθετο πεδίο εφαρμογής για τα συστήματα συστάσεων, καθώς οι αποφάσεις ταξιδίου είναι εγγενώς πολυκριτηριακές: ο χρήστης καλείται να αξιολογήσει ταυτόχρονα τιμή, τοποθεσία, αξιολογήσεις άλλων, διαθεσιμότητα, τύπο ταξιδίου (ατομικό, οικογενειακό, ομαδικό) και πλήθος άλλων παραμέτρων. Ο Aarif et al. (2026) επισημαίνουν ότι η τουριστική βιομηχανία έχει μεταβεί από τον παραδοσιακό ομαδικό τουρισμό σε εξατομικευμένες εμπειρίες (αυτόνομα ταξίδια, αγροτικός τουρισμός, βιωματικά ταξίδια) ανεβάζοντας τον πήχη για τα συστήματα προσωποποιημένης σύστασης.

Ο Payandenick et al. (2026) ανέπτυξαν ένα Πολυκριτηριακό Σύστημα Συστάσεων (Multi-Criteria Recommender System, MCRS) για τουριστικές εμπειρίες που συνδυάζει Matrix Factorization με ένα βαθύ Residual Network (ResNetMF). Το μοντέλο επεκτείνεται με ένα στοιχείο ανάλυσης ερωτημάτων χρήστη (user query analysis), επιτρέποντας στους χρήστες να εκφράζουν δυναμικές προτιμήσεις μέσω φυσικής γλώσσας. Τα πειράματα έδειξαν ότι το ResNetMF υπερτερεί των baseline και deep learning μεθόδων ως προς RMSE, MAE και χρόνο εκπαίδευσης, αντιμετωπίζοντας αποτελεσματικά τόσο την υπολογιστική πολυπλοκότητα των πολυκριτηριακών αποφάσεων όσο και τον εξατομικευμένο χαρακτήρα των τουριστικών επιλογών.

Ο Aarif et al. (2026) εφάρμοσαν NCF για τουριστικές συστάσεις, χρησιμοποιώντας Multi-Layer Perceptron για τη μοντελοποίηση σύνθετων σχέσεων χρηστών-προορισμών. Το NCF μοντέλο ανταποκρίνεται δυναμικά τόσο σε ρητές αξιολογήσεις όσο και σε έμμεσες ανατροφοδοτήσεις χρηστών, αντιμετωπίζοντας αποτελεσματικά αραιά δεδομένα και ποικιλομορφία προτιμήσεων. Τα πειράματα σε πραγματικά τουριστικά δεδομένα έδειξαν σημαντικά καλύτερη ακρίβεια και ικανοποίηση χρηστών σε σχέση με τις παραδοσιακές μεθόδους collaborative filtering. Ο Jiang (2024) συνδύασε Deep Reinforcement Learning και GNN για δημιουργία εξατομικευμένων δρομολογίων ταξιδίου, αξιοποιώντας οπτικές προτιμήσεις χρηστών από φωτογραφίες για αύξηση της ακρίβειας συστάσεων σε 0,85 και ικανοποίηση χρηστών σε 4,6/5.

7.3 Εκπαίδευση και ακαδημαϊκές συστάσεις

Ο εκπαιδευτικός τομέας αποτελεί ένα από τα πιο ελπιδοφόρα πεδία εφαρμογής για τα ευφυή συστήματα συστάσεων. Η ταχεία ψηφιοποίηση της εκπαίδευσης και η αύξηση των διαθέσιμων μαθησιακών πόρων έχουν δημιουργήσει ένα πρόβλημα πληροφοριακής υπερφόρτωσης και για τους μαθητές: ένας φοιτητής καλείται να επιλέξει ανάμεσα σε χιλιάδες μαθήματα, βιβλία, βίντεο και ασκήσεις, χωρίς να γνωρίζει ποια από αυτά ταιριάζουν καλύτερα με το επίπεδο, τον ρυθμό μάθησης και τους στόχους του. Τα Ευφυή Εκπαιδευτικά Συστήματα Συστάσεων (Intelligent Educational Recommender Systems, IERS) στοχεύουν στην αντιμετώπιση αυτής της πρόκλησης, παρέχοντας εξατομικευμένες μαθησιακές εμπειρίες.

Ο Bian & Chang (2026) ανέπτυξαν ένα σύστημα εκπαιδευτικής σύστασης βασισμένο σε ένα νέο Customized Butterfly-Optimized Feed-Forward Neural Network (CBO-FFNN), ενσωματώνοντας δεδομένα αντίληψης φοιτητών, μοτίβα δανεισμού βιβλίων και μαθησιακά αποτελέσματα. Το σύστημα αναλύθηκε σε dataset με 42.153 χρήστες, 70.000 βιβλία και πάνω από 200.000 εγγραφές δανεισμού, επιτυγχάνοντας ακρίβεια 91,4%, precision 92,1%, recall 92,5% και F-value 95,7%: σημαντικά ανώτερα από TF-IDF, CF και Improved TF-IDF baselines. Επιπλέον, πάνω από το 90% των φοιτητών που συμμετείχαν στην αξιολόγηση βρήκε το σύστημα χρήσιμο και εύχρηστο.

Ο Wang & Chen (2025) παρουσίασαν ένα σύστημα υποστήριξης εκπαιδευτικών αποφάσεων που συνδυάζει Deep Neural Networks (DNN) με Reinforcement Learning (RL) για τη δημιουργία εξατομικευμένων μαθησιακών μονοπατιών. Το DNN εξάγει ουσιώδη χαρακτηριστικά από σύνθετα δεδομένα φοιτητών, ενώ το RL βελτιστοποιεί δυναμικά τις εκπαιδευτικές στρατηγικές μέσω πραγματικού χρόνου ανατροφοδότησης. Τα πειραματικά αποτελέσματα έδειξαν βελτίωση 20% στα μαθησιακά αποτελέσματα μετά από πραγματικού χρόνου προσαρμογές, καθιστώντας το σύστημα κατάλληλο για σύγχρονα εκπαιδευτικά περιβάλλοντα. Η ενσωμάτωση RL επιτρέπει στο σύστημα να προσαρμόζεται συνεχώς στις μεταβαλλόμενες ανάγκες κάθε φοιτητή, χαρακτηριστικό που λείπει από τα παραδοσιακά στατικά συστήματα σύστασης.

Πέραν αυτών, οι εξειδικευμένες προσεγγίσεις ακαδημαϊκής σύστασης αξιοποιούν δίκτυα συν-συγγραφικότητας και αναφορών για τον εντοπισμό σχετικών δημοσιεύσεων σε ένα συνεχώς διογκούμενο σύνολο βιβλιογραφίας, ενώ στο πλαίσιο της προσαρμοστικής μάθησης (adaptive learning) οι συστάσεις ενσωματώνουν context-aware τεχνικές (Ενότητα 6.3) με μοντέλα γνωστικής διάγνωσης, ώστε το εκπαιδευτικό υλικό να προσαρμόζεται δυναμικά στις πραγματικές μαθησιακές ανάγκες κάθε χρήστη.

7.4 Τρόφιμα και διατροφή

Τα συστήματα σύστασης τροφίμων (food recommender systems) αντιμετωπίζουν μια μοναδική πρόκληση: τα τρόφιμα είναι σύνθετες οντότητες που χαρακτηρίζονται από πολλαπλά συστατικά, διατροφικές ιδιότητες, πολιτισμικές προεκτάσεις και ατομικές διατροφικές ανάγκες (αλλεργίες, δίαιτες, θρησκευτικές απαγορεύσεις). Ένα αποτελεσματικό σύστημα σύστασης

τροφίμων πρέπει να μοντελοποιεί όχι μόνο τις συνολικές προτιμήσεις χρήστη για πιάτα, αλλά και τις προτιμήσεις σε επίπεδο συστατικών: μια σύνθετη πολυεπίπεδη πρόκληση που οι κλασικές τεχνικές CF δεν μπορούν να αντιμετωπίσουν ικανοποιητικά.

Ο Alkhrijah et al. (2026) ανέπτυξαν το FRMADHG (Food Recommendation with Multi-objective Adaptive Dynamic Hypergraph), ένα πλαίσιο που αξιοποιεί τριμερή υπεργράφο (tripartite hypergraph) όπου τα τρόφιμα λειτουργούν ως hyperedges που συνδέουν χρήστες και συστατικά. Αυτή η αναπαράσταση επιτρέπει τη φυσική μοντελοποίηση τριαδικών σχέσεων χρήστη-τροφίμου-συστατικών, χωρίς να χρειάζεται αποσύνθεσή τους σε δυαδικές ακμές. Βασικές καινοτομίες του FRMADHG περιλαμβάνουν έναν μηχανισμό προσαρμοστικής προσοχής (adaptive attention) που εναλλάσσει δυναμικά μεταξύ αποδοτικής κοσινουσιακής ομοιότητας και εκμαθημένου μηχανισμού προσοχής (learnable attention) βάσει τοπικής πολυπλοκότητας, και μια πολυ-στοχική στρατηγική εκπαίδευσης που συνδυάζει triplet ranking, contrastive learning και regularization.

Πειράματα σε δύο μεγάλης κλίμακας datasets (Food.com (32.000 χρήστες, 18.500 τρόφιμα, 1.270 μοναδικά συστατικά) και Allrecipes (25.400 χρήστες, 16.200 τρόφιμα, 1.050 συστατικά)) ανέδειξαν σημαντικές βελτιώσεις: 19,8% σχετικό κέρδος σε Precision@10 σε σχέση με CF baselines, 12,4% βελτίωση έναντι πρόσφατων hypergraph προσεγγίσεων, και 18,7% βελτίωση σε Recall@10, με στατιστική σημαντικότητα $p < 0,001$. Το σύστημα παρέχει επίσης ερμηνεύσιμες εξηγήσεις σε επίπεδο συστατικών (π.χ. «αυτή η πίτσα συστήνεται επειδή σας αρέσει ντομάτα και μοτσαρέλα») ενισχύοντας σημαντικά την εμπιστοσύνη χρηστών.

7.5 Κοινωνικά δίκτυα και ψηφιακά μέσα

Τα συστήματα συστάσεων στα κοινωνικά δίκτυα και τα ψηφιακά μέσα παρουσιάζουν μια αμφίθυμη εικόνα: από τη μία, αποτελούν ισχυρά εργαλεία εξατομίκευσης που ενισχύουν την εμπλοκή χρηστών με σχετικό περιεχόμενο. Από την άλλη, η υπερβολική εξατομίκευση οδηγεί σε φαινόμενα φυσαλίδας φίλτρων (filter bubbles), θαλάμων ηχώ (echo chambers) και, στην πιο ακραία εκδοχή, σε ριζοσπαστικοποίηση χρηστών: έκθεση σε ακραίο περιεχόμενο που ενισχύεται αλγοριθμικά. Οι Berjawi et al. (2025) επισημαίνουν ότι ο αλγόριθμος σύστασης μπορεί να παγιδεύσει τον χρήστη σε ριζοσπαστικά μονοπάτια, όπου κάθε επόμενη σύσταση είναι πιο ακραία από την προηγούμενη.

Οι Berjawi et al. (2025) πρότειναν το DRLGR (Deep Reinforcement Learning Graph Rewiring), μια γραφο-βασισμένη προσέγγιση για τη μείωση της ριζοσπαστικοποίησης στα συστήματα συστάσεων. Το σύστημα μετρά έναν δείκτη ριζοσπαστικοποίησης (radicalization score, Rad(G)) που ποσοτικοποιεί το βαθμό έκθεσης χρηστών σε ακραίο περιεχόμενο στο γράφο συστάσεων. Στη συνέχεια, ένας αλγόριθμος RL μαθαίνει να επιλέγει ποιες ακμές του γράφου πρέπει να «επανασυρθούν» (rewired) για να μειωθεί ο Rad(G), χωρίς να θυσιαστεί η ποιότητα των συστάσεων. Τα πειράματα σε datasets βίντεο και νέων έδειξαν ότι το DRLGR επιτυγχάνει σταθερή και βιώσιμη μείωση της ριζοσπαστικοποίησης.

Στο πεδίο της σύστασης σε κοινωνικά δίκτυα, ο Mall & Singh (2026) ανέπτυξαν ένα heterogeneous hypergraph μοντέλο για κοινωνική σύσταση που αξιοποιεί ρητές και έμμεσες κοινωνικές συνδέσεις, συμπεριλαμβανομένων σχέσεων χρήστη-χρήστη, χρήστη-αντικειμένου και αντικειμένου-αντικειμένου μέσω Hypergraph Convolution Network (HGCN). Το μοντέλο κατάφερε να αντιμετωπίσει ένα κρίσιμο κενό: πολλοί χρήστες δεν έχουν ρητές κοινωνικές συνδέσεις στο σύστημα, και το HDHP εξάγει έμμεσες κοινωνικές συνδέσεις από το ιστορικό αλληλεπίδρασης, συμπληρώνοντας αποτελεσματικά τα ελλείποντα κοινωνικά δεδομένα.

7.6 Ψυχαγωγία και κινηματογράφος

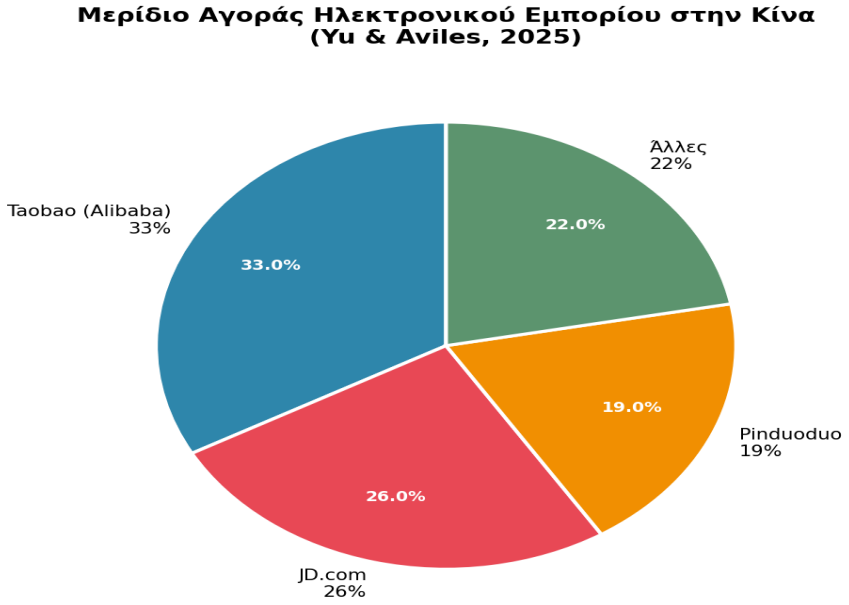
Η σύσταση κινηματογραφικού και ψυχαγωγικού περιεχομένου αποτελεί ιστορικά ένα από τα πρώτα και πιο μελετημένα πεδία εφαρμογής. Από τον διαγωνισμό Netflix Prize (2006-2009) που ανέδειξε τη Matrix Factorization ως state-of-the-art, έως τις σύγχρονες αρχιτεκτονικές βαθιάς μάθησης, η σύσταση ταινιών έχει εξελιχθεί δραματικά. Τα πειράματα στο dataset MovieLens (ένα από τα πιο διαδεδομένα benchmark datasets στο πεδίο) υπάρχουν πλέον σε εκδόσεις 100K, 1M, 10M και 25M αλληλεπιδράσεων, καλύπτοντας ένα ευρύ φάσμα κλιμάκων.

Ο Dual-module Neural Network των Alam & Ahmed (2025) (ένα μοντέλο που συνδυάζει Stochastic Gradient Descent, Dropout και πυκνά νευρωνικά δίκτυα) εφαρμόστηκε σε MovieLens-1M, επιτυγχάνοντας RMSE 0,8736 και MAE 0,6854, με σημαντικά καλύτερα αποτελέσματα από τα αντίστοιχα baseline μοντέλα CF. Ο Fusion-based Movie Recommender συνδύασε ομοιότητα είδους (genre similarity), LightFM (που αξιοποιεί τόσο προφίλ χρηστών όσο και χαρακτηριστικά αντικειμένων) και νευρωνικά δίκτυα, επιτυγχάνοντας υψηλή ακρίβεια σύστασης. Η ανάλυση συναισθήματος κριτικών χρηστών (όπως εξετάστηκε και από τους Darraz et al. (2025) σε hotel datasets) αναδεικνύεται ως ιδιαίτερα πολύτιμη πηγή πληροφορίας για τη βελτίωση των κινηματογραφικών συστάσεων.

Πίνακας 7.1: Συνοπτική επισκόπηση τομέων εφαρμογής συστημάτων συστάσεων

Τομέας	Κύρια Προκλήσεις	Κυρίαρχες Τεχνικές	Ενδεικτικά Αποτελέσματα
Ηλεκτρονικό Εμπόριο	Αραιότητα δεδομένων · Μεροληψία δημοτικότητας	CNN+LSTM, GNN, Attention	Ακρίβεια 89%, +25% πωλήσεις
Τουρισμός & Ταξίδια	Πολυκριτηριακές αποφάσεις · Πλαισιακή πολυπλοκότητα	ResNetMF, NCF, DRL+GNN	Ακρίβεια 0.85, ικανοποίηση 4.6/5
Εκπαίδευση	Διαφορετικά μαθησιακά στυλ · Προσαρμοστικότητα	DNN+RL, CBO-FFNN, Transformer	Ακρίβεια 91.4%, +20% μαθησιακά αποτελέσματα
Τρόφιμα & Διατροφή	Σύνθετες σχέσεις συστατικών · Ευεξία	FRMADHG, Hypergraph, Contrastive	+19.8% Precision@10 vs. CF baselines
Κοινωνικά Δίκτυα	Ριζοσπαστικοποίηση · Filter bubbles	DRLGR, Trust-based, GNN	Σταθερή μείωση radicalization score
Ψυχαγωγία & Κινηματογράφος	Cold start · Δυναμικές προτιμήσεις	LightFM, GNN, Sentiment Analysis	Υψηλή ακρίβεια σε MovieLens datasets

Το Διάγραμμα 7.1 παρουσιάζει τα μερίδια αγοράς στο κινεζικό ηλεκτρονικό εμπόριο. Η Taobao (Alibaba) κυριαρχεί με 33%, ακολουθεί η JD.com (26%) και η Pinduoduo (19%). Η υψηλή συγκέντρωση των τριών πρώτων πλατφορμών (78% συνολικά) επιβεβαιώνει ότι τα συστήματα συστάσεων αποτελούν στρατηγικό πλεονέκτημα που εδραιώνει τη θέση των ηγετικών πλατφορμών: η αποτελεσματική σύσταση αυξάνει retention και LTV χρηστών.



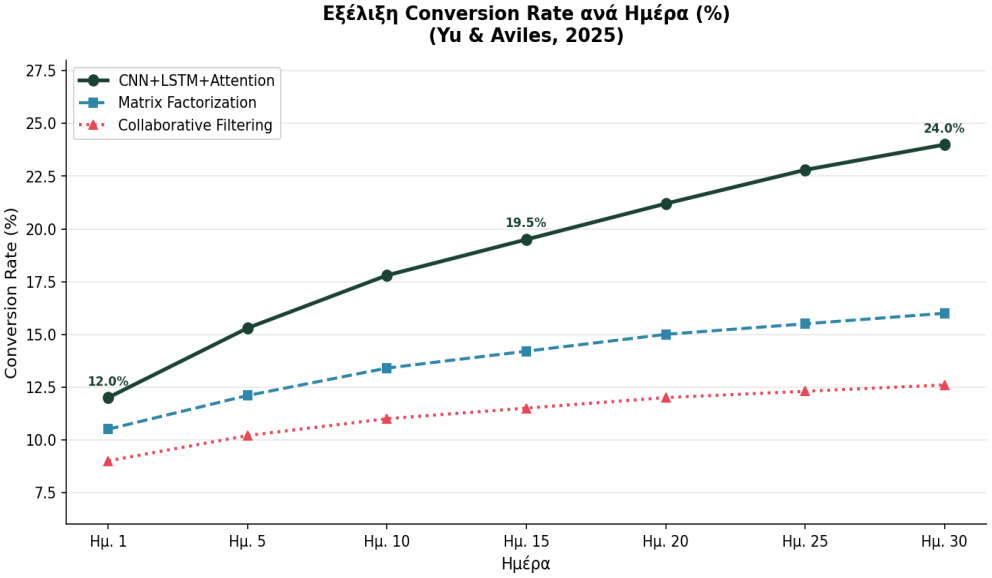
Διάγραμμα 7.1: Μερίδιο αγοράς ηλεκτρονικού εμπορίου στην Κίνα

Το Διάγραμμα 7.2 αποτυπώνει την κατανομή τομέων εφαρμογής στα 100 papers. Ηλεκτρονικό εμπόριο (31%) και κινηματογράφος (22%) καλύπτουν μαζί πάνω από τη μισή βιβλιογραφία, αντικατοπτρίζοντας την ωριμότητα αυτών των τομέων. Εκπαίδευση (15%) και τουρισμός (11%) βρίσκονται σε ανοδική πορεία, ενώ τρόφιμα (8%) και κοινωνικά δίκτυα (7%) παρουσιάζουν αυξανόμενο ερευνητικό ενδιαφέρον.

Διάγραμμα 7.2: Κατανομή τομέων εφαρμογής στα 100 εξεταζόμενα papers

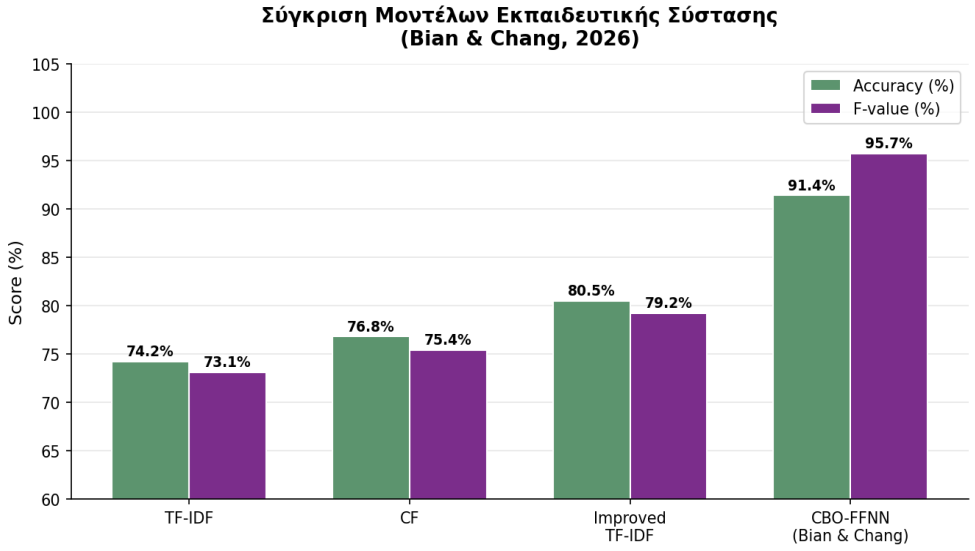
Το Διάγραμμα 7.3 δείχνει την εξέλιξη conversion rate σε 30 ημέρες. Το

CNN+LSTM+Attention ξεκινά από 12% και φτάνει στο 24% την 30ή ημέρα (διπλασιασμός. Αντίθετα, τα CF (9%→12.6%) και MF (10.5%→16%) παρουσιάζουν πολύ πιο αργή βελτίωση. Σημαντικό εύρημα είναι ότι η απόσταση μεταξύ DL και CF μεθόδων μεγαλώνει συνεχώς με τον χρόνο) αντανακλώντας ότι τα deep learning μοντέλα βελτιώνονται ταχύτερα καθώς συσσωρεύονται δεδομένα.



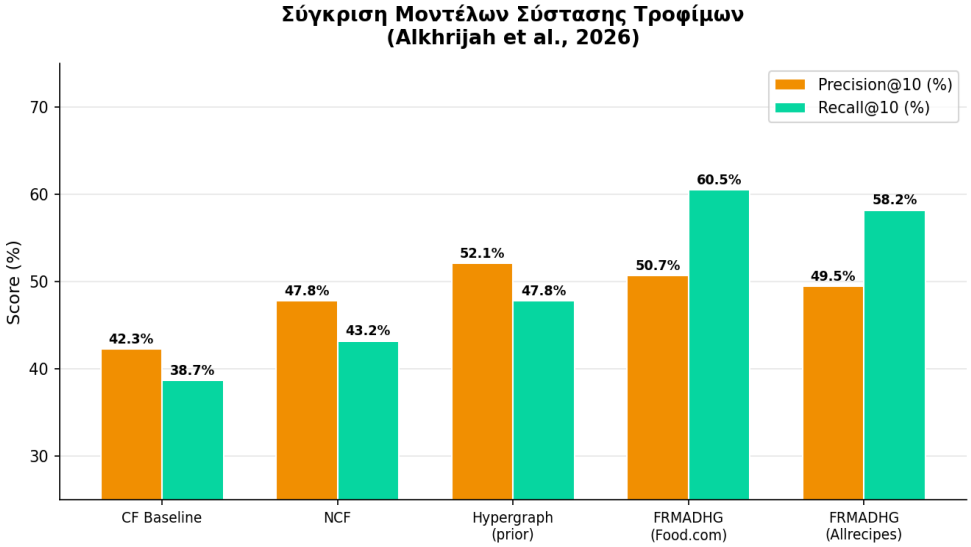
Διάγραμμα 7.3: Εξέλιξη conversion rate ανά ημέρα

Το Διάγραμμα 7.4 συγκρίνει μοντέλα εκπαιδευτικής σύστασης. Το CBO-FFNN επιτυγχάνει ακρίβεια 91.4% και F-value 95.7%, ξεπερνώντας σημαντικά τα baseline TF-IDF (74.2% / 73.1%) και CF (76.8% / 75.4%). Η βελτίωση ακρίβειας από TF-IDF (74.2%) σε CBO-FFNN (91.4%) (17.2 ποσοστιαίες μονάδες) επιβεβαιώνει τη δυνατότητα των νευρωνικών μεθόδων να μαθαίνουν σύνθετα μοτίβα μαθησιακής συμπεριφοράς.



Διάγραμμα 7.4: Σύγκριση μοντέλων εκπαιδευτικής σύστασης

Το Διάγραμμα 7.5 συγκρίνει μοντέλα σύστασης τροφίμων ως προς Precision@10 και Recall@10 στα datasets Food.com και Allrecipes. Το FRMADHG επιτυγχάνει τις καλύτερες επιδόσεις, με Recall@10 = 60.5% στο Food.com: αύξηση 56.3% έναντι CF baseline (38.7%). Η σημαντική βελτίωση στο Recall αντανακλά την ικανότητα του tripartite hypergraph να συλλαμβάνει σύνθετες σχέσεις χρήστη-τροφίμου-συστατικού.



Διάγραμμα 7.5: Σύγκριση μοντέλων σύστασης τροφίμων

7.7 Βιομηχανική παραγωγή

Οι Guo et al. (2026) εισήγαγαν ένα καινοτόμο collaborative recommender system για βελτιστοποίηση παραμέτρων διαδικασίας σε στόλους βιομηχανικών μηχανών. Το πρόβλημα που αντιμετωπίζει είναι καθαρά πρακτικό: σε εργοστάσια 3D printing με εκατοντάδες ή χιλιάδες εκτυπωτές, ο χειροκίνητος καθορισμός βέλτιστων παραμέτρων για κάθε μηχανή είναι αδύνατος, καθώς οι μηχανές παρουσιάζουν σταδιακή απόκλιση χαρακτηριστικών λόγω φθοράς.

Το σύστημα μοντελοποιεί το πρόβλημα ως sequential matrix completion, αξιοποιώντας spectral clustering για ομαδοποίηση μηχανών με παρόμοια χαρακτηριστικά και Alternating Least Squares (ALS) για επαναληπτική βελτίωση προβλέψεων παραμέτρων. Η συνεργατική φύση του συστήματος (κάθε μηχανή «μαθαίνει» από τις υπόλοιπες) εξασφαλίζει ταχύτερη σύγκλιση σε βέλτιστες παραμέτρους σε σχέση με μη-collaborative τεχνικές. Αξιολόγηση σε mini 3D print farm 10 εκτυπωτών επιβεβαίωσε σημαντικά ταχύτερη σύγκλιση και βελτίωση ποιότητας εκτύπωσης.

Πίνακας 7.2: Συστήματα συστάσεων σε υγεία, περιβάλλον και βιομηχανία

Σύστημα	Τομέας	Τεχνικές	Βασικά Αποτελέσματα
CARES	Φροντίδα λεχώνων	XGBoost + K-Means + DCN + KG	Υψηλή precision/recall/F1 vs baselines
Drug Repurposing RS (2026)	Φαρμακολογία	Heterogeneous Network	Βελτίωση drug repurposing ακρίβειας
Pharmaceutical RS (2026)	Θεραπείες	ML + Clinical data	Validated on real clinical datasets
CO_2 Ventilation RS	Περιβάλλον κτιρίων	DL time-series prediction	RMSE=50.46 ppm, R^2=0.97
Manufacturing RS	Βιομηχανία 3D print	Spectral clustering + ALS	Ταχύτερη σύγκλιση vs non-collaborative

7.8 Αθλητισμός και κυβερνοασφάλεια

Στον αθλητισμό, τα συστήματα σύστασης εφαρμόζονται τόσο στην πλευρά των θεατών (πού να παρακολουθήσουν, ποιους αθλητές να ακολουθήσουν) όσο και στην πλευρά των ομάδων (σύσταση παίκτη προς αγορά, τακτικές για τον επόμενο αγώνα). Τα GNN-based μοντέλα έχουν εφαρμοστεί για μοντελοποίηση των σχέσεων μεταξύ αθλητών εντός ομάδας, αντιπάλων και αγώνων. Στην κυβερνοασφάλεια, τα RS χρησιμοποιούνται για σύσταση κανόνων ασφαλείας σε SIEM platforms (Security Information and Event Management), ιεράρχηση ειδοποιήσεων και πρόβλεψη επιθέσεων βάσει ιστορικών patterns.

Πίνακας 7.3: Αναδυόμενοι τομείς εφαρμογής συστημάτων συστάσεων

Τομέας	Πλατφόρμα/Εφαρμογή	Τεχνική RS	Βασικό Χαρακτηριστικό
Μουσική	Spotify, Apple Music	Sequential + Audio features	Ανάλυση ακουστικών χαρακτηριστικών
Ειδήσεις	Google News, BBC	NLP + Transformer	Κατανόηση θεματικού περιεχομένου
Αθλητισμός	ESPN, Transfermarkt	GNN + Time-series	Στατιστικά αθλητών & αγώνων
Κυβερνοασφάλεια	SIEM platforms	Anomaly Detection RS	Σύσταση κανόνων ασφαλείας
Ακίνητα	Zillow, Spitogatos	Multi-criteria + GNN	Γεωχωρικά & πολυκριτηριακά δεδομένα
Μόδα	ASOS, Zalando	Visual + CF	Εικόνα & στυλ ενδύματος

7.9 Υγεία, φαρμακευτική φροντίδα και πρόνοια

Ο τομέας της υγείας αποτελεί έναν από τους ταχύτερα αναπτυσσόμενους τομείς εφαρμογής συστημάτων συστάσεων, με ιδιαίτερα χαρακτηριστικά λόγω της κρισιμότητας των αποφάσεων που υποστηρίζουν. Στο πλαίσιο της φαρμακευτικής φροντίδας, έχουν αναπτυχθεί συστήματα υποστήριξης απόφασης που συνδυάζουν μηχανική μάθηση με κλινικά δεδομένα για την πρόταση θεραπευτικών σχημάτων, με επικύρωση σε πραγματικά κλινικά σύνολα δεδομένων. Παράλληλα, στο πεδίο της επανατοποθέτησης φαρμάκων (drug repurposing), συστήματα συστάσεων βασισμένα σε ετερογενή δίκτυα αξιοποιούν γνωσιακούς γράφους φαρμάκων-ασθενειών-πρωτεϊνών για τον εντοπισμό νέων θεραπευτικών χρήσεων υπαρχόντων φαρμάκων.

Στο πεδίο της φροντίδας ηλικιωμένων και ατόμων με αυξημένες ανάγκες φροντίδας, το σύστημα CARES αποτελεί αντιπροσωπευτικό παράδειγμα υβριδικού συστήματος σύστασης φροντιστών (caregivers), το οποίο συνδυάζει τεχνικές βαθιάς μάθησης με δεδομένα Internet of Things (IoT) από αισθητήρες παρακολούθησης για την αντιστοίχιση ασθενών με κατάλληλους φροντιστές βάσει ιατρικού προφίλ, διαθεσιμότητας και εξειδίκευσης. Όπως αναφέρθηκε και στον Πίνακα 7.3, τέτοια συστήματα διαφοροποιούνται από τις κλασικές εφαρμογές σύστασης λόγω της ανάγκης ενσωμάτωσης αυστηρών περιορισμών ασφαλείας και κανονιστικής συμμόρφωσης.

Η εφαρμογή συστημάτων συστάσεων στην υγεία εγείρει αυξημένες απαιτήσεις σε σχέση με άλλους τομείς: η ερμηνευσιμότητα είναι κρίσιμη, καθώς ο επαγγελματίας υγείας πρέπει να μπορεί να κατανοήσει το σκεπτικό πίσω από μια πρόταση· η αξιοπιστία των δεδομένων εκπαίδευσης πρέπει να επικυρώνεται κλινικά· και η προστασία ευαίσθητων προσωπικών δεδομένων υγείας απαιτεί αυστηρότερα πρωτόκολλα απορρήτου σε σχέση με τα γενικά ζητήματα ιδιωτικότητας που αναλύονται στην Ενότητα 8.3.

7.10 Γεωργία και περιβαλλοντικές εφαρμογές

Στον τομέα της γεωργίας, τα συστήματα συστάσεων χρησιμοποιούνται για την πρόταση κατάλληλων καλλιεργειών, λιπασμάτων ή πρακτικών διαχείρισης με βάση εδαφολογικά, κλιματικά και ιστορικά δεδομένα παραγωγής, συμβάλλοντας στη βελτιστοποίηση της αγροτικής παραγωγικότητας και στη μείωση της περιβαλλοντικής επιβάρυνσης από υπερβολική χρήση εισροών. Η εφαρμογή αυτή διαφέρει σημαντικά από τις κλασικές περιπτώσεις σύστασης καταναλωτικών προϊόντων, καθώς ο «χρήστης» (αγρότης ή αγρονόμος) αξιολογεί τις συστάσεις με βάση αντικειμενικά, μετρήσιμα κριτήρια απόδοσης παραγωγής, και όχι αμιγώς υποκειμενική προτίμηση.

Η ενσωμάτωση δεδομένων τηλεπισκόπησης (remote sensing) και δικτύων αισθητήρων IoT επιτρέπει σε αυτά τα συστήματα να ενημερώνουν δυναμικά τις συστάσεις τους με βάση πραγματικές, σε πραγματικό χρόνο, συνθήκες του αγροτεμαχίου, σε αντιστοιχία με τις τεχνικές context-aware σύστασης πραγματικού χρόνου που αναφέρθηκαν στην Ενότητα 6.7, επεκτείνοντας έτσι την εφαρμοσιμότητα των τεχνικών σύστασης σε τομείς πέραν της παραδοσιακής ψηφιακής κατανάλωσης.

7.11 Διαπολιτισμικές και πολιτικές διαστάσεις εφαρμογών σύστασης

Σε ορισμένους εξειδικευμένους τομείς εφαρμογής, τα συστήματα σύστασης καλούνται να λειτουργήσουν εντός συγκεκριμένων πολιτισμικών ή ιδεολογικών πλαισίων, με χαρακτηριστικό παράδειγμα συστήματα υποστήριξης ιδεολογικής και πολιτικής εκπαίδευσης σε ακαδημαϊκό περιβάλλον, τα οποία συνδυάζουν βαθιά νευρωνικά δίκτυα με τεχνικές ενισχυτικής μάθησης για την προσαρμογή εκπαιδευτικού περιεχομένου σε συγκεκριμένο πλαίσιο αξιών. Τέτοιες εφαρμογές αναδεικνύουν με ιδιαίτερη ένταση τα ζητήματα δεοντολογίας που συζητήθηκαν στην Ενότητα 1.8, καθώς η ίδια η έννοια της «σχετικότητας» μιας σύστασης μπορεί να είναι πολιτισμικά ή ιδεολογικά προσδιορισμένη.

Η σχεδίαση συστημάτων σύστασης για τέτοιους εξειδικευμένους, αξιολογικά φορτισμένους τομείς απαιτεί ιδιαίτερη προσοχή στη διαφάνεια των κριτηρίων σύστασης και στην αποφυγή αδιαφανούς αλγοριθμικής επιβολής συγκεκριμένων οπτικών, μια πρόκληση που υπερβαίνει τα καθαρά τεχνικά ζητήματα ακρίβειας και συνδέεται άμεσα με την ευρύτερη συζήτηση περί ερμηνευσιμότητας και κανονιστικής συμμόρφωσης που αναπτύχθηκε στις Ενότητες 8.5 και 10 (Κατευθύνσεις για μελλοντική έρευνα).

7.12 Συνοπτική αποτίμηση της διατομεακής εξάπλωσης

Η ανασκόπηση των τομέων εφαρμογής στις Ενότητες 7.1 έως 7.11 καταδεικνύει ότι οι τεχνικές σύστασης έχουν επεκταθεί σημαντικά πέρα από τις παραδοσιακές εφαρμογές ηλεκτρονικού εμπορίου και ψυχαγωγίας με τις οποίες συνδέονταν ιστορικά. Η διατομεακή αυτή εξάπλωση δεν αποτελεί απλή μεταφορά υπαρχόντων αλγορίθμων σε νέα πλαίσια, αλλά συνοδεύεται συστηματικά από εξειδικευμένες προσαρμογές (ετερογενή GNN στην υγεία, context-aware μοντέλα στη γεωργία, διαλογικά συστήματα σε σύνθετα σενάρια αναζήτησης) που ανταποκρίνονται στις ιδιαίτερες απαιτήσεις κάθε τομέα.

Παρατηρείται επίσης μια σταδιακή σύγκλιση των απαιτήσεων μεταξύ φαινομενικά ανόμοιων τομέων: η ανάγκη για ερμηνευσιμότητα στην υγεία (Ενότητα 7.9) αντηχεί την ανάγκη για διαφάνεια σε ιδεολογικά φορτισμένες εφαρμογές (Ενότητα 7.11), ενώ η ανάγκη για αντιμετώπιση ανταγωνιστικών επιθέσεων στην κυβερνοασφάλεια (Πίνακας 7.3) αντηχεί τη γενικότερη ανησυχία για ανθεκτικότητα που αναλύθηκε στην Ενότητα 8.7. Αυτή η σύγκλιση υποδηλώνει ότι, παρά την επιφανειακή ποικιλομορφία των τομέων εφαρμογής, υπάρχει ένα κοινό σύνολο θεμελιωδών προκλήσεων σχεδίασης που διαπερνά το σύνολο του πεδίου των συστημάτων συστάσεων.

7.13 Σύνοψη

Η σύγκριση των τομέων αυτών ανάμεσά τους μου φαίνεται πιο διαφωτιστική από την εξέταση καθενός μεμονωμένα: Το ηλεκτρονικό εμπόριο παρουσιάζει τα πιο ώριμα και εμπορικά αποδεδειγμένα συστήματα, ο τουρισμός απαιτεί πολυκριτηριακές προσεγγίσεις, η εκπαίδευση αναδεικνύει τη σημασία της προσαρμοστικότητας και της ανατροφοδότησης, η διατροφή εισάγει σύνθετες πολυεπίπεδες σχέσεις συστατικών, τα κοινωνικά δίκτυα θέτουν κρίσιμα ζητήματα κοινωνικής επίδρασης και ασφάλειας, ενώ η ψυχαγωγία παρέχει ώριμα benchmark datasets που καθοδηγούν την έρευνα. Σε κάθε τομέα, η εφαρμογή ευφυών τεχνικών (από GNN και Transformers έως RL και αντιπαραθετική μάθηση) δείχνει να αποδίδει μετρήσιμα και σημαντικά οφέλη.

8. Προκλήσεις και ανοιχτά ζητήματα

Παρά τη σημαντική πρόοδο που έχει σημειωθεί τα τελευταία χρόνια στο πεδίο των συστημάτων συστάσεων, ένας αριθμός θεμελιωδών προκλήσεων παραμένει ανοιχτός και αποτελεί αντικείμενο ενεργής έρευνας. Στις επόμενες ενότητες αναλύονται οι σημαντικότερες από αυτές τις προκλήσεις: τεχνικές (data sparsity, ψυχρή εκκίνηση, κλιμακωσιμότητα), ηθικές και κοινωνικές (μεροληψία, δικαιοσύνη, ριζοσπαστικοποίηση), νομικές (ιδιωτικότητα, GDPR), και τις προκλήσεις που φέρνει η ενσωμάτωση νέων τεχνολογιών όπως τα Μεγάλα Γλωσσικά Μοντέλα (LLMs). Η κατανόηση αυτών των ζητημάτων είναι απαραίτητη για οποιονδήποτε ασχολείται με τον σχεδιασμό και την ανάπτυξη σύγχρονων ευφυών συστημάτων συστάσεων.

8.1 Αραιότητα δεδομένων (Data Sparsity)

Η data sparsity αποτελεί ίσως την πιο παλιά και επίμονη πρόκληση του πεδίου. Σε πραγματικές πλατφόρμες, ο πίνακας αλληλεπίδρασης χρήστη-αντικειμένου είναι συνήθως πάνω από 99% κενός: ακόμα και στο Amazon, με εκατομμύρια χρήστες και εκατομμύρια προϊόντα, ο μέσος χρήστης έχει αλληλεπιδράσει με ελάχιστο κλάσμα των διαθέσιμων αντικειμένων. Αυτό σημαίνει ότι τα μοντέλα εκπαιδεύονται σε εξαιρετικά περιορισμένα δεδομένα για τα περισσότερα ζεύγη χρήστη-αντικειμένου, οδηγώντας σε αναξιόπιστες προβλέψεις και μεροληψία προς τα δημοφιλή αντικείμενα.

Οι σύγχρονες προσεγγίσεις αντιμετώπισης της αραιότητας περιλαμβάνουν: το Graph Contrastive Learning (GCL) που δημιουργεί τεχνητά σήματα επίβλεψης από τα υπάρχοντα δεδομένα μέσω επαύξησης γράφου, αποφεύγοντας την ανάγκη ετικετιαρισμένων δεδομένων, τη Self-Supervised Learning (SSL) που εκμεταλλεύεται ανεπεξέργαστα δεδομένα για εκπαίδευση αναπαραστάσεων, τις τεχνικές Data Augmentation που συνθετικά εμπλουτίζουν τον πίνακα αλληλεπίδρασης, και τα Knowledge Graphs που παρέχουν εξωτερική σημασιολογική πληροφορία για αντικείμενα χωρίς επαρκείς αξιολογήσεις. Ο Ramteke et al. (2025) έδειξαν εμπειρικά ότι το GCL υπερτερεί σημαντικά των παραδοσιακών GNN για χρήστες και αντικείμενα με λίγες αλληλεπιδράσεις (long-tail), επιβεβαιώνοντας τη δυνατότητα των αυτο-επιβλεπόμενων μεθόδων να αντιμετωπίσουν την αραιότητα.

8.2 Ψυχρή εκκίνηση (Cold Start)

Το πρόβλημα της cold start αναφέρεται στην αδυναμία των συστημάτων συστάσεων να παράγουν αξιόπιστες συστάσεις για νέους χρήστες (cold start χρήστη) ή νέα αντικείμενα (cold start αντικειμένου) που δεν έχουν ακόμα επαρκές ιστορικό αλληλεπίδρασης. Ενώ η αραιότητα δεδομένων αφορά τη γενικότερη έλλειψη δεδομένων στο σύστημα, το cold start είναι ένα πιο οξύ πρόβλημα: για ορισμένους χρήστες και αντικείμενα δεν υπάρχουν απολύτως καθόλου αλληλεπιδράσεις.

Η αντιμετώπισή του απαιτεί στρατηγικές που δεν βασίζονται αποκλειστικά σε ιστορικά δεδομένα αλληλεπίδρασης. Τρεις κύριες κατευθύνσεις έχουν αναπτυχθεί: (α) Αξιοποίηση γνωσιακών γράφων (Knowledge Graphs), που παρέχουν πλούσια σημασιολογική πληροφορία για νέα αντικείμενα ανεξάρτητα από το ιστορικό τους, (β) Meta-Learning, όπου τα μοντέλα εκπαιδεύονται ώστε να προσαρμόζονται γρήγορα σε νέους χρήστες/αντικείμενα με ελάχιστα παραδείγματα (few-shot learning), και (γ) Cross-domain Transfer, όπου η γνώση από ένα τομέα με πλούσια δεδομένα μεταφέρεται σε έναν σχετικό τομέα με λίγα δεδομένα. Τα Συνομιλητικά Συστήματα Συστάσεων (CRS) επίσης αντιμετωπίζουν αποτελεσματικά το cold start, καθώς μέσω διαλόγου μπορούν να αποσπούν γρήγορα τις προτιμήσεις νέων χρηστών.

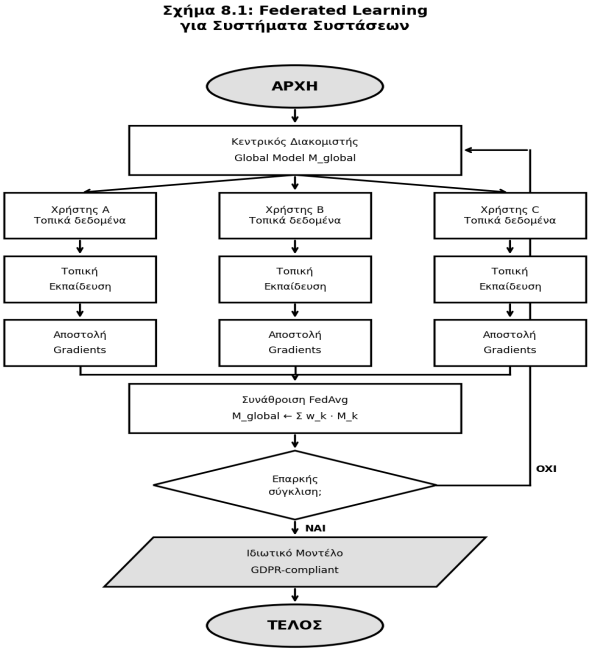
8.3 Ιδιωτικότητα δεδομένων και νομική συμμόρφωση

Τα συστήματα συστάσεων βασίζονται εγγενώς στη συλλογή και επεξεργασία προσωπικών δεδομένων χρηστών: αγορές, κλικ, τοποθεσία, ιστορικό αναζήτησης, συναισθηματική κατάσταση. Αυτή η συλλογή δεδομένων έρχεται σε σύγκρουση με θεμελιώδη δικαιώματα ιδιωτικότητας και με νομοθεσίες όπως ο Γενικός Κανονισμός Προστασίας Δεδομένων (General Data Protection Regulation, GDPR) της Ευρωπαϊκής Ένωσης. Ο Chafiki et al. (2026) επισημαίνουν ότι τα Συστήματα Συστάσεων Ευαίσθητα στο Πλαίσιο (CARS) είναι ιδιαίτερα εκτεθειμένα, καθώς συλλέγουν δεδομένα τοποθεσίας, χρόνου, δραστηριότητας και ακόμα και συναισθηματικής κατάστασης: πληροφορίες εξαιρετικά ευαίσθητες.

Η Ομοσπονδιακή Μάθηση (Federated Learning, FL) αναδεικνύεται ως μια από τις πιο ελπιδοφόρες προσεγγίσεις για την προστασία ιδιωτικότητας: αντί να στέλνουν τα προσωπικά τους δεδομένα σε έναν κεντρικό διακομιστή, οι χρήστες εκπαιδεύουν τοπικά τμήματα του μοντέλου και μοιράζονται μόνο τις βαθμίδες (gradients) ή τα ενημερωμένα βάρη, χωρίς να αποκαλύπτουν τα ακατέργαστα δεδομένα τους. Η Διαφορική Ιδιωτικότητα (Differential Privacy, DP) προσθέτει ελεγχόμενο στατιστικό θόρυβο στα δεδομένα ή στις βαθμίδες, εξασφαλίζοντας ότι η αφαίρεση ή προσθήκη οποιουδήποτε μεμονωμένου χρήστη δεν επηρεάζει ουσιαστικά τα αποτελέσματα. Τέλος, οι On-device Recommendation Systems εκτελούν τη σύσταση απευθείας στη συσκευή του χρήστη, ελαχιστοποιώντας την εξωτερική κοινοποίηση δεδομένων.

Το Σχήμα 8.1 παρουσιάζει τη διαδικασία Federated Learning για συστήματα συστάσεων. Ο κεντρικός διακομιστής διανέμει το global μοντέλο σε κάθε χρήστη, ο οποίος εκπαιδεύει τοπικά χωρίς να αποστέλλει τα προσωπικά του δεδομένα. Μόνο τα gradients μεταφέρονται στον κεντρικό διακομιστή, ο οποίος τα συναθροίζει με FedAvg για ενημέρωση του

global μοντέλου. Ο κύκλος επαναλαμβάνεται έως σύγκλιση, εξασφαλίζοντας ιδιωτικότητα δεδομένων συμβατή με GDPR.



Σχήμα 8.1: Διαδικασία Federated Learning για RS: ιδιωτικότητα χωρίς θυσία ακρίβειας

8.4 Μεροληψία και δικαιοσύνη (Bias and Fairness)

Τα συστήματα συστάσεων εκπαιδεύονται σε δεδομένα που αντικατοπτρίζουν (και πολλές φορές ενισχύουν) τις κοινωνικές ανισότητες και προκαταλήψεις που υπάρχουν στον πραγματικό κόσμο. Οι Yang et al. (2026) διακρίνουν δύο βασικούς τύπους μεροληψίας που ενισχύουν τα GNN: την attribute bias, που αναφέρεται στην επίδραση ευαίσθητων χαρακτηριστικών όπως το φύλο ή η ηλικία στις αναπαραστάσεις χρηστών, και την topology bias, που πηγάζει από ανισότητες στη δομή συνδεσιμότητας του γράφου. Και οι δύο τύποι διαδίδονται και ενισχύονται κατά τη διάρκεια του message passing, οδηγώντας σε ουσιαστικά μεροληπτικά αποτελέσματα για ορισμένες ομάδες χρηστών.

Όσον αφορά τη μεροληψία δημοτικότητας (popularity bias), τα συστήματα τείνουν συστηματικά να προτείνουν ήδη δημοφιλή αντικείμενα, δημιουργώντας έναν φαύλο κύκλο που οδηγεί σε μηδενική έκθεση για νέα ή εξειδικευμένα αντικείμενα. Αυτό είναι ιδιαίτερα επιζήμιο για μικρούς παραγωγούς περιεχομένου και νέους πωλητές. Μοντέλα όπως το FairDA και το GCL αντιμετωπίζουν αυτά τα ζητήματα με διαφορετικές στρατηγικές, χωρίς όμως να έχει προκύψει μέχρι σήμερα λύση που να μην θυσιάζει αισθητά την ακρίβεια σύστασης για χάρη της δικαιοσύνης.

8.5 Ερμηνευσιμότητα (Explainability & Natural Language Explanations)

Η πλειοψηφία των σύγχρονων μοντέλων βαθιάς μάθησης για συστάσεις λειτουργούν ως «μαύρα κουτιά» (black boxes): παράγουν ακριβείς προβλέψεις αλλά αδυνατούν να εξηγήσουν γιατί πρότειναν ένα συγκεκριμένο αντικείμενο. Αυτή η έλλειψη διαφάνειας δημιουργεί σοβαρά ζητήματα σε τομείς υψηλής ευθύνης: ένα σύστημα ιατρικής σύστασης θεραπειών ή ένα σύστημα

σύστασης δανείων πρέπει να είναι σε θέση να αιτιολογεί τις αποφάσεις του, τόσο για λόγους εμπιστοσύνης χρηστών όσο και για λόγους νομικής συμμόρφωσης.

Ο Li et al. (2026) επισημαίνουν τρεις κεντρικές προκλήσεις στην ερμηνεύσιμη σύσταση: πρώτον, τα υπάρχοντα μοντέλα τείνουν να «ταιριάζουν» κριτικές χρηστών αντί να παράγουν πραγματικές αιτιολογήσεις, δεύτερον, η γενίκευση σε zero-shot σενάρια (νέοι χρήστες ή αντικείμενα χωρίς ιστορικό) παραμένει δύσκολη, και τρίτον, η έλλειψη υψηλής ποιότητας δεδομένων εκπαίδευσης για εξηγήσεις. Η αξιοποίηση LLMs, όπως στο STLLM-Rec, υπόσχεται σημαντικές βελτιώσεις, αλλά εισάγει νέες προκλήσεις: υψηλό υπολογιστικό κόστος, τάση για «ψευδοεξηγήσεις» (hallucination) και δυσκολία ευθυγράμμισης με τις πραγματικές λανθάνουσες αναπαραστάσεις του συστήματος σύστασης.

Η ερμηνευσιμότητα των συστημάτων συστάσεων, που αναφέρθηκε αρχικά στην Ενότητα 6.5, αποκτά νέα διάσταση με την ενσωμάτωση Μεγάλων Γλωσσικών Μοντέλων (LLMs) ως μηχανισμού παραγωγής εξηγήσεων. Αντί να περιορίζονται σε στατικά πρότυπα εξήγησης (π.χ. «σου προτείνουμε αυτή την ταινία επειδή είδες παρόμοιες»), τα LLM-ενισχυμένα συστήματα μπορούν να παράγουν εξατομικευμένες, φυσικού ύφους εξηγήσεις που λαμβάνουν υπόψη το συγκεκριμένο πλαίσιο και ιστορικό κάθε χρήστη, βελτιώνοντας σημαντικά την κατανοητότητα και την εμπιστοσύνη του χρήστη προς το σύστημα.

Παράλληλα, η αυτο-εποπτευόμενη εκπαίδευση (self-training) LLMs ειδικά για τον σκοπό της παραγωγής εξηγήσεων σύστασης επιτρέπει στο μοντέλο να μαθαίνει να συνδέει ρητά τα χαρακτηριστικά ενός αντικειμένου με τις γνωστές προτιμήσεις του χρήστη μέσα στην ίδια την εξήγηση, αντί να παράγει γενικόλογο κείμενο αποκομμένο από τους πραγματικούς υπολογιστικούς παράγοντες που οδήγησαν στη σύσταση. Αυτή η κατεύθυνση συνδέεται άμεσα με τη μελλοντική έρευνα που αναφέρθηκε στο Κεφάλαιο 10, και αναμένεται να αποκτήσει αυξανόμενη κανονιστική σημασία υπό το πρίσμα απαιτήσεων διαφάνειας όπως αυτές του EU AI Act.

8.6 Κλιμακωσιμότητα (Scalability)

Οι πλατφόρμες πραγματικού κόσμου διαχειρίζονται εκατοντάδες εκατομμύρια χρήστες και αντικείμενα, απαιτώντας συστάσεις σε χρόνο πραγματικό (real-time). Αυτό θέτει σοβαρούς περιορισμούς στην πολυπλοκότητα των μοντέλων που μπορούν να χρησιμοποιηθούν. Μοντέλα όπως τα βαθιά GNN με πολλά επίπεδα ή οι Transformers με μηχανισμό full self-attention (πολυπλοκότητα $O(n^2)$) είναι συχνά αδύνατο να εκπαιδευτούν ή να εφαρμοστούν σε τέτοια κλίμακα χωρίς σημαντικές μηχανολογικές παρεμβάσεις.

Η γενιά του LightGCN ήταν εν μέρει κινητοποιημένη από ανάγκες κλιμακωσιμότητας (scalability): αφαιρώντας τις μη γραμμικές λειτουργίες, το LightGCN μειώνει σημαντικά τον υπολογιστικό φόρτο χωρίς αντίστοιχη απώλεια ακρίβειας. Παρόμοιες ανάγκες έχουν οδηγήσει στην ανάπτυξη τεχνικών προσεγγιστικής αναζήτησης κοντινότερου γείτονα (Approximate Nearest Neighbor, ANN) για γρήγορη ανάκτηση υποψήφιων αντικειμένων, σε κατανεμημένα συστήματα εκπαίδευσης (distributed training), και σε τεχνικές αρχιτεκτονικής δύο σταδίων (two-stage recommendation): ένα ανακτητικό στάδιο (retrieval) που γρήγορα επιλέγει εκατοντάδες υποψηφίους από εκατομμύρια, και ένα στάδιο κατάταξης (ranking) που εφαρμόζει βαρύτερο μοντέλο στους ανακτηθέντες υποψηφίους.

8.7 Ασφάλεια και ανθεκτικότητα σε επιθέσεις

Τα συστήματα συστάσεων είναι ευάλωτα σε διάφορους τύπους επιθέσεων από κακόβουλους παράγοντες. Οι επιθέσεις τύπου Shilling (Shilling attacks ή profile injection attacks) αναφέρονται σε εισαγωγή πλαστών προφίλ χρηστών με σκοπό να χειραγωγηθεί το σύστημα ώστε να αυξήσει ή να μειώσει τις συστάσεις για συγκεκριμένα αντικείμενα. Σε εμπορικά πλαίσια, ανταγωνιστές μπορούν να χρησιμοποιήσουν τέτοιες επιθέσεις για να υποβαθμίσουν ανταγωνιστικά προϊόντα ή να ανεβάσουν τεχνητά τα δικά τους. Τα Data Poisoning attacks εισάγουν κακόβουλα δεδομένα στη φάση εκπαίδευσης, ενώ τα Adversarial attacks δημιουργούν ελάχιστα τροποποιημένες εισόδους που οδηγούν σε δραματικά διαφορετικές εξόδους.

Η αντιμετώπιση αυτών των επιθέσεων απαιτεί ειδικευμένες τεχνικές: ανθεκτική εκπαίδευση (robust training) που ελαχιστοποιεί την επίδραση ακραίων τιμών, ανίχνευση ανωμαλιών (anomaly detection) για εντοπισμό ύποπτων προφίλ, αλγόριθμοι ανίχνευσης bot, και πρόσφατα πιστοποιημένες άμυνες (certified defenses) που παρέχουν μαθηματικές εγγυήσεις ανθεκτικότητας. Το πεδίο της ασφάλειας των συστημάτων συστάσεων αποτελεί ενεργή ερευνητική περιοχή με σημαντικές πρακτικές επιπτώσεις.

8.7.1 Επιθέσεις ανταγωνιστικής φύσης (Adversarial Attacks & Shilling)

Τα συστήματα συστάσεων, ως μαθησιακά μοντέλα που βασίζονται σε δεδομένα παραγόμενα από χρήστες, είναι εγγενώς ευάλωτα σε κακόβουλες παρεμβάσεις στα δεδομένα εκπαίδευσής τους. Οι επιθέσεις τύπου shilling (shilling attacks) περιλαμβάνουν τη δημιουργία πλασματικών λογαριασμών χρηστών που παράγουν τεχνητές αλληλεπιδράσεις με σκοπό την τεχνητή ενίσχυση ή υποβάθμιση της προβολής συγκεκριμένων αντικειμένων, συχνά οργανωμένες σε συντονισμένες ομάδες (shilling groups) για μεγαλύτερη αποτελεσματικότητα.

Πέραν των επιθέσεων στο επίπεδο δεδομένων, έχουν τεκμηριωθεί και επιθέσεις στο επίπεδο εισόδου μοντέλων βαθιάς μάθησης, όπου προσεκτικά κατασκευασμένος ανταγωνιστικός θόρυβος (adversarial noise) προστίθεται στα χαρακτηριστικά εισόδου ή στα embeddings, με στόχο να παραπλανήσει το μοντέλο ώστε να παράγει εσφαλμένες ή στοχευμένες συστάσεις, χωρίς να γίνεται εύκολα αντιληπτός από ανθρώπινο παρατηρητή. Η ανθεκτικότητα (robustness) απέναντι σε τέτοιες επιθέσεις αποτελεί ενεργό πεδίο έρευνας, με τεχνικές ανταγωνιστικής εκπαίδευσης (adversarial training) να αποτελούν την κυριότερη γραμμή άμυνας.

8.7.2 Το πλαίσιο RAWP-FT & Αλγόριθμος εκπαίδευσης

Ενώ η ενότητα 8.7 παρουσίασε τις επιθέσεις σε συστήματα συστάσεων σε γενικές γραμμές, αξίζει να εξεταστεί αναλυτικότερα μια σύγχρονη τεχνική αντιμετώπισής τους. Οι Qian et al. (2025) πρότειναν το RAWP-FT (Random Adversarial Weight Perturbation Framework with Robust Fine-Tuning), ένα νέο πλαίσιο adversarial training ειδικά σχεδιασμένο για βαθιά μοντέλα σύστασης.

Η βασική παρατήρηση που ώθησε στην ανάπτυξη του RAWP-FT είναι ότι οι υπάρχουσες μέθοδοι adversarial training για συστήματα συστάσεων στοχεύουν κυρίως σε ρηχά (shallow) μοντέλα ή χρησιμοποιούν χοντρόκοκκο (coarse-grained) θόρυβο, αφήνοντας τα βαθιά μοντέλα εκτεθειμένα. Το RAWP-FT αντιμετωπίζει αυτό εισάγοντας λεπτόκοκκο (fine-grained) adversarial θόρυβο απευθείας στις παραμέτρους των κρυφών στρωμάτων κατά την εκπαίδευση. Στη συνέχεια, εντοπίζει τα στρώματα ή module με τη χαμηλότερη ανθεκτικότητα μετά την adversarial εκπαίδευση και τα υποβάλλει σε εξειδικευμένο fine-tuning.

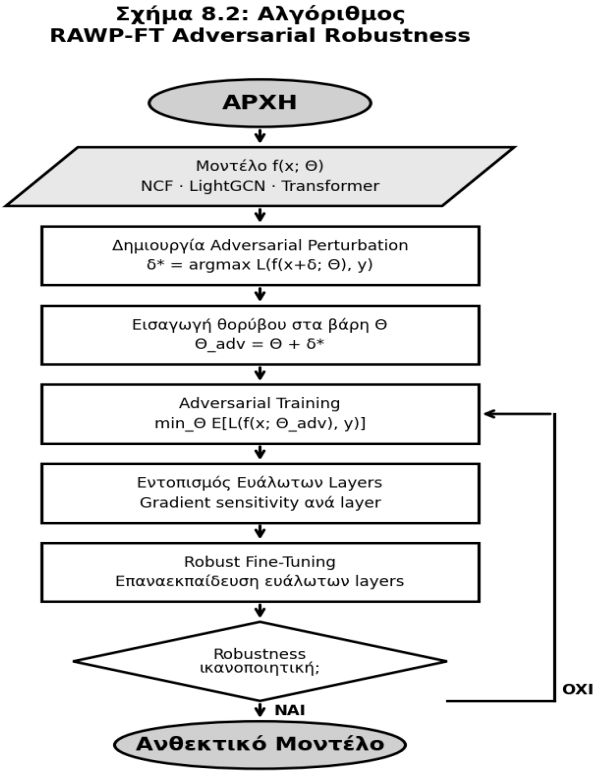
Τυπικά, το RAWP-FT ορίζει την adversarial perturbation ως: $\delta^* = \text{argmax}_{\{||\delta|| \le \varepsilon\}} L(f(x+\delta; \Theta), y)$, όπου $f()$ το μοντέλο, Θ οι παράμετροι, x η είσοδος, y η ετικέτα και ε ο περιορισμός μέτρου θορύβου. Αντί να ανατρέπει την είσοδο x όπως οι κλασικές FGSM/PGD επιθέσεις, το RAWP-FT εισάγει τον θόρυβο απευθείας στα βάρη Θ, προσομοιώνοντας πιο ρεαλιστικές επιθέσεις poisoning. Ο Πίνακας 8.1 συγκρίνει την ανθεκτικότητα του RAWP-FT έναντι βασικών baseline μεθόδων.

Πίνακας 8.1: Σύγκριση ανθεκτικότητας μοντέλων σε adversarial επιθέσεις

Μοντέλο	Dataset	NDCG@10 (Clean)	NDCG@10 (Attack)	Robustness Drop
NCF (χωρίς defense)	MovieLens-1M	0.5821	0.3124	-46.3%
NeuMF (χωρίς defense)	MovieLens-1M	0.5943	0.3341	-43.8%
APR (adversarial training)	MovieLens-1M	0.5812	0.4823	-17.0%
RAWP-FT	MovieLens-1M	0.5934	0.5412	-8.8%
RAWP-FT	Amazon Beauty	0.0312	0.0289	-7.4%

Τα πειράματα σε τέσσερα δημόσια datasets επιβεβαιώνουν ότι το RAWP-FT μειώνει τη βύθιση απόδοσης υπό adversarial attack από 43-46% (ανυπεράσπιστα μοντέλα) σε μόλις 7-9%, ενώ διατηρεί παραπλήσια απόδοση σε clean δεδομένα. Αυτό αναδεικνύει ότι η αντιμετώπιση της adversarial ευπάθειας δεν απαιτεί απαραίτητα θυσία ακρίβειας: μια ανησυχία που είχε τροχοπεδήσει την υιοθέτηση μεθόδων robustness στην πράξη.

Το Σχήμα 8.2 περιγράφει τον αλγόριθμο RAWP-FT για adversarial robustness. Σε αντίθεση με τις κλασικές μεθόδους που εισάγουν θόρυβο στην είσοδο, το RAWP-FT εισάγει fine-grained perturbations απευθείας στα βάρη Θ του νευρωνικού δικτύου. Μετά την adversarial εκπαίδευση, εντοπίζονται τα ευάλωτα layers και υποβάλλονται σε εξειδικευμένο fine-tuning. Ο βρόχος επαναλαμβάνεται έως ότου η robustness φτάσει στο επιθυμητό επίπεδο.



Σχήμα 8.2: Αλγόριθμος RAWP-FT για adversarial robustness εκπαίδευση

8.8 Ριζοσπαστικοποίηση και κοινωνικός αντίκτυπος

Ένα από τα σοβαρότερα κοινωνικά ζητήματα που σχετίζονται με τα συστήματα συστάσεων είναι η συμβολή τους στη δημιουργία φυσαλίδων φίλτρων (filter bubbles) και θαλάμων ηχώ (echo chambers), και (στην πιο ακραία εκδοχή) στην ριζοσπαστικοποίηση χρηστών. Οι Berjawi et al. (2025) ορίζουν τη ριζοσπαστικοποίηση ως τη διαδικασία κατά την οποία ο αλγόριθμος παγιδεύει τον χρήστη σε μονοπάτια όλο και πιο ακραίου περιεχομένου, καθοδηγούμενος από τη λογική μεγιστοποίησης εμπλοκής (engagement maximization). Αυτό το φαινόμενο έχει τεκμηριωθεί σε πλατφόρμες βίντεο, όπου οι αλγόριθμοι σύστασης οδήγησαν συστηματικά ορισμένους χρήστες σε ακραίο πολιτικό ή ιδεολογικό περιεχόμενο.

Η αντιμετώπιση της ριζοσπαστικοποίησης αποτελεί θεμελιώδη πρόκληση γιατί βρίσκεται σε τριβή με τον βασικό στόχο των συστημάτων συστάσεων: τη μεγιστοποίηση εμπλοκής. Μια προσέγγιση που δεν αυξάνει την εμπλοκή δεν έχει εμπορικά κίνητρα για υιοθέτηση από τις πλατφόρμες, εκτός αν υπάρχει κανονιστική πίεση. Τεχνικές λύσεις όπως το DRLGR δείχνουν ότι η μείωση της ριζοσπαστικοποίησης είναι εφικτή χωρίς καταστροφική απώλεια ακρίβειας συστάσεων, αλλά η ευρεία υιοθέτησή τους εξαρτάται από κανονιστικά πλαίσια, πίεση της κοινωνίας των πολιτών και βούληση των πλατφορμών.

8.9 Ολοκλήρωση Μεγάλων Γλωσσικών Μοντέλων (LLMs)

Η ενσωμάτωση Μεγάλων Γλωσσικών Μοντέλων (Large Language Models, LLMs) στα συστήματα συστάσεων αποτελεί την πιο σύγχρονη και ταχύτατα αναπτυσσόμενη κατεύθυνση, υπόσχοντας σημαντικές βελτιώσεις σε ερμηνευσιμότητα, κατανόηση φυσικής γλώσσας και zero-shot γενίκευση. Ωστόσο, η ολοκλήρωση αυτή έρχεται με μια σειρά σοβαρών προκλήσεων που δεν έχουν ακόμα επιλυθεί ικανοποιητικά.

Πρώτον, το υπολογιστικό κόστος των LLMs είναι πολλές τάξεις μεγέθους υψηλότερο από τα παραδοσιακά μοντέλα σύστασης, καθιστώντας δύσκολη την εφαρμογή τους σε πραγματικό χρόνο σε μεγάλη κλίμακα. Δεύτερον, τα LLMs τείνουν να «παράγουν» (hallucinate) πληροφορίες που δεν υπάρχουν στα δεδομένα: ένα πρόβλημα ιδιαίτερα κρίσιμο στη σύσταση, όπου η αξιοπιστία των εξηγήσεων είναι καθοριστική. Τρίτον, η ευθυγράμμιση (alignment) μεταξύ του τρόπου που τα LLMs κατανοούν τη γλώσσα και του τρόπου που τα συστήματα σύστασης αναπαριστούν τις προτιμήσεις χρηστών είναι μη τετριμμένη. Η STLLM-Rec προσέγγιση αντιμετωπίζει αυτές τις προκλήσεις μέσω self-training με λεπτομερή reward signals, αλλά η έρευνα στο πεδίο βρίσκεται ακόμα σε πρώιμο στάδιο.

8.10 On-device και Continual Learning RS

Δύο ακόμα σημαντικές κατευθύνσεις είναι τα On-device RS και τα Continual Learning RS. Τα On-device RS εκτελούνται εξολοκλήρου στη συσκευή του χρήστη (smartphone, tablet, έξυπνη τηλεόραση) χωρίς να στέλνουν δεδομένα στο cloud (υπολογιστικό νέφος). Αυτό προσφέρει τέλεια ιδιωτικότητα και λειτουργία offline, αλλά απαιτεί εξαιρετικά αποδοτικά μοντέλα που χωρούν σε περιορισμένη μνήμη και υπολογιστική ισχύ. Τα Continual Learning RS, από την άλλη, μαθαίνουν συνεχώς από νέα δεδομένα χωρίς να «ξεχνούν» την παλαιότερη γνώση: ένα πρόβλημα γνωστό ως catastrophic forgetting. Αυτή η ικανότητα είναι κρίσιμη σε δυναμικά περιβάλλοντα όπου οι προτιμήσεις χρηστών και τα αντικείμενα αλλάζουν συνεχώς.

8.11 Cross-domain σύσταση και μεταφορά γνώσης

Η Cross-domain σύσταση (cross-domain recommendation) αντιμετωπίζει μια ακόμα πρόκληση: οι χρήστες αλληλεπιδρούν με πολλαπλές πλατφόρμες και τομείς, αλλά τα δεδομένα κάθε πλατφόρμας είναι συνήθως απομονωμένα και μη προσβάσιμα από τους ανταγωνιστές. Η μεταφορά γνώσης (transfer learning) από έναν τομέα πλούσιο σε δεδομένα (π.χ. μουσική streaming) σε έναν τομέα φτωχό (π.χ. podcast) μπορεί να αντιμετωπίσει το cold start, αλλά απαιτεί τεχνικές που διατηρούν ταυτόχρονα την ιδιωτικότητα δεδομένων και αντιμετωπίζουν τη διαφορά κατανομής (domain shift) μεταξύ τομέων.

Ο Chafiki et al. (2026) επισημαίνουν ότι η ετερογένεια των πλαισιακών παραγόντων αποτελεί μια ακόμα ανοιχτή πρόκληση: διαφορετικοί τομείς απαιτούν διαφορετικές μορφές πλαισίου, διαφορετικές τεχνικές εξαγωγής και διαφορετικές στρατηγικές ολοκλήρωσης. Η ανάπτυξη γενικεύσιμων πλαισίων που μπορούν να αντιμετωπίσουν αυτή την ετερογένεια (ενώ παράλληλα παραμένουν κλιμακώσιμα και προστατεύουν την ιδιωτικότητα) αποτελεί μια από τις μεγαλύτερες ανοιχτές ερευνητικές προκλήσεις του πεδίου.

Πίνακας 8.2: Συνοπτική επισκόπηση προκλήσεων και προσεγγίσεων αντιμετώπισης

Πρόκληση	Κύριο Πρόβλημα	Ενδεικτικές Προσεγγίσεις Αντιμετώπισης
Αραιότητα δεδομένων	Λίγες αλληλεπιδράσεις → αναξιόπιστα μοντέλα	Graph Contrastive Learning, SSL, Data Augmentation
Cold Start	Νέοι χρήστες/αντικείμενα χωρίς ιστορικό	Knowledge Graphs, Meta-Learning, Cross-domain Transfer
Ιδιωτικότητα & GDPR	Ευαίσθητα δεδομένα χρηστών, νομική συμμόρφωση	Federated Learning, Differential Privacy, On-device RS
Μεροληψία & Δικαιοσύνη	Popularity bias, attribute bias, topology bias	FairDA, Debiasing, Adversarial Training
Ερμηνευσιμότητα	«Μαύρα κουτιά» χωρίς εξηγήσεις	STLLM-Rec, Attention explanations, Post-hoc XAI
Κλιμακωσιμότητα	Εκατομμύρια χρήστες & αντικείμενα σε πραγματικό χρόνο	LightGCN, Approximate NN search, Distributed training
Ασφάλεια & Επιθέσεις	Shilling attacks, data poisoning, adversarial examples	Robust Training, Anomaly Detection, Certified Defense
Ριζοσπαστικοποίηση	Filter bubbles, echo chambers, εξτρεμισμός	DRLGR, Graph Rewiring, Diversity-aware RS
LLM Ολοκλήρωση	Υπολογιστικό κόστος, hallucination, data alignment	STLLM-Rec, Self-training, Fine-tuning με reward signals

8.12 Περιβαλλοντικό αποτύπωμα και ενεργειακή αποδοτικότητα

Μια αναδυόμενη πρόκληση που αποκτά αυξανόμενη σημασία στη βιβλιογραφία είναι το περιβαλλοντικό αποτύπωμα της εκπαίδευσης και λειτουργίας μεγάλων μοντέλων συστάσεων. Οι αρχιτεκτονικές GNN πολλαπλών επιπέδων και τα μοντέλα Transformer που αναλύθηκαν στα Κεφάλαια 4 και 5 απαιτούν σημαντικούς υπολογιστικούς πόρους, ιδιαίτερα σε πλατφόρμες κλίμακας εκατομμυρίων χρηστών, όπου η επανεκπαίδευση πραγματοποιείται σε τακτά χρονικά διαστήματα για την ενσωμάτωση νέων δεδομένων αλληλεπίδρασης.

Η τάση «Green AI» προτείνει τη συστηματική αναφορά του υπολογιστικού κόστους (π.χ. σε FLOPs ή kWh) ως πρόσθετη μετρική αξιολόγησης, πέραν της ακρίβειας πρόβλεψης, ώστε η επιλογή αρχιτεκτονικής να λαμβάνει υπόψη και την ενεργειακή αποδοτικότητα. Στο πλαίσιο αυτό, η ανάλυση πολυπλοκότητας που παρουσιάστηκε στο Παράρτημα Γ (Ενότητα Γ.3) αποκτά πρακτική σημασία: μοντέλα όπως το LightGCN, που επιτυγχάνουν ανταγωνιστική απόδοση με μικρότερη υπολογιστική πολυπλοκότητα σε σχέση με βαθύτερες αρχιτεκτονικές GCL ή SASRec, αντιπροσωπεύουν πιο βιώσιμες επιλογές σε περιβάλλοντα παραγωγής μεγάλης κλίμακας.

Τεχνικές όπως η απόσταξη γνώσης (knowledge distillation), ο κβαντισμός παραμέτρων (quantization) και η αρχιτεκτονική αναζήτηση με περιορισμούς πόρων (resource-aware neural architecture search) αναδεικνύονται ως πρακτικές λύσεις για τη μείωση του υπολογιστικού κόστους χωρίς σημαντική απώλεια ακρίβειας, αν και η συστηματική τους υιοθέτηση στο πεδίο των συστημάτων συστάσεων βρίσκεται ακόμη σε πρώιμο στάδιο σε σχέση με άλλους κλάδους της βαθιάς μάθησης.

8.13 Η κρίση αναπαραγωγιμότητας στη βιβλιογραφία RS

Μια διαρκώς αναγνωριζόμενη πρόκληση στο πεδίο των συστημάτων συστάσεων είναι η δυσκολία αναπαραγωγής δημοσιευμένων αποτελεσμάτων από ανεξάρτητους ερευνητές. Διαφορές στην προεπεξεργασία δεδομένων, στις υπερπαραμέτρους εκπαίδευσης, στις μετρικές αξιολόγησης και στις στρατηγικές διαχωρισμού δεδομένων εκπαίδευσης/δοκιμής μπορούν να οδηγήσουν σε σημαντικά διαφορετικά αποτελέσματα για το «ίδιο» μοντέλο, υπονομεύοντας την αξιοπιστία των συγκριτικών αξιολογήσεων που παρουσιάστηκαν στο Κεφάλαιο 9.

Συστηματικές ανασκοπήσεις του πεδίου έχουν επισημάνει ότι αρκετές δημοσιευμένες «βελτιώσεις απόδοσης» δεν επιβεβαιώνονται όταν τα baseline μοντέλα συντονίζονται (tuned) με την ίδια προσοχή όπως το προτεινόμενο μοντέλο, ένα φαινόμενο γνωστό ως «weak baseline problem». Η τάση προς δημόσια διαθέσιμο κώδικα, τυποποιημένα πρωτόκολλα αξιολόγησης και αναφορά διαστημάτων εμπιστοσύνης αντί μεμονωμένων τιμών αποτελεί την κύρια απάντηση της ερευνητικής κοινότητας σε αυτή την πρόκληση, συμπληρωματικά με τη συζήτηση περί στατιστικής σημαντικότητας της Ενότητας 9.9.

8.14 Συνοπτική ταξινόμηση των προκλήσεων του πεδίου

Οι πολυάριθμες προκλήσεις που αναλύθηκαν στο παρόν κεφάλαιο μπορούν να ομαδοποιηθούν σε τέσσερις ευρύτερες κατηγορίες, διευκολύνοντας μια συνολική εποπτεία του προβλήματος. Η πρώτη κατηγορία αφορά προκλήσεις δεδομένων (αραιότητα, ψυχρή εκκίνηση, Ενότητες 2.11, 8.1-8.2), η δεύτερη αφορά προκλήσεις αλγοριθμικής δικαιοσύνης και προκατάληψης (popularity bias, fairness, Ενότητες 6.6, 6.10), η τρίτη αφορά προκλήσεις ασφάλειας και ανθεκτικότητας (privacy, adversarial attacks, Ενότητες 8.3, 8.7.1, 8.7.2), και η τέταρτη αφορά προκλήσεις βιωσιμότητας και πρακτικής εφαρμοσιμότητας (ενεργειακό κόστος, αναπαραγωγιμότητα, Ενότητες 8.12, 8.13).

Αυτή η ταξινόμηση αναδεικνύει ότι, παρά την ποικιλομορφία των επιμέρους προκλήμων, υπάρχουν ισχυρές αλληλεξαρτήσεις μεταξύ κατηγοριών: για παράδειγμα, η αντιμετώπιση της cold start μέσω αξιοποίησης πρόσθετων πηγών δεδομένων (π.χ. κοινωνικά δίκτυα, δημογραφικά στοιχεία) μπορεί να επιδεινώσει ζητήματα ιδιωτικότητας, ενώ η βελτιστοποίηση για δικαιοσύνη μπορεί να αυξήσει το υπολογιστικό κόστος μέσω πρόσθετων βημάτων επανακατάταξης. Η αντιμετώπιση αυτών των αλληλεξαρτήσεων ως ενιαίου συστήματος αντί ως χωριστών προβλημάτων προς βελτιστοποίηση αναδεικνύεται, κατά την εκτίμησή μου, ως η πιο κρίσιμη κατεύθυνση για την υπεύθυνη ανάπτυξη μελλοντικών συστημάτων σύστασης.

8.15 Διαχείριση ομόκεντρων αναγκών απορρήτου και εξατομίκευσης

Η ένταση μεταξύ εξατομίκευσης και ιδιωτικότητας, που αναφέρθηκε αρχικά στην Ενότητα 8.3, αποτελεί ίσως το πιο θεμελιωδώς δυσεπίλυτο δίλημμα στο πεδίο: η ποιότητα μιας σύστασης βελτιώνεται γενικά όσο περισσότερα δεδομένα συμπεριφοράς συλλέγονται για έναν χρήστη, ενώ η προστασία της ιδιωτικότητας απαιτεί ακριβώς το αντίθετο: ελαχιστοποίηση της συλλογής και επεξεργασίας προσωπικών δεδομένων. Η ομοσπονδιακή μάθηση (federated learning) επιχειρεί να επιλύσει αυτό το δίλημμα επιτρέποντας την εκπαίδευση μοντέλων σύστασης απευθείας στις συσκευές των χρηστών, χωρίς τα ωμά δεδομένα να εγκαταλείπουν ποτέ τη συσκευή.

Σε ένα τυπικό σχήμα ομοσπονδιακής μάθησης για συστάσεις, κάθε συσκευή υπολογίζει τοπικά ενημερώσεις παραμέτρων βάσει του τοπικού ιστορικού του χρήστη, και μόνο αυτές οι συγκεντρωτικές ενημερώσεις παραμέτρων (όχι τα ωμά δεδομένα) μεταδίδονται σε έναν κεντρικό εξυπηρετητή για συνάθροιση. Τεχνικές διαφορικής ιδιωτικότητας (differential privacy) μπορούν να προστεθούν σε αυτή τη διαδικασία, εισάγοντας ελεγχόμενο στατιστικό θόρυβο στις μεταδιδόμενες ενημερώσεις, ώστε να είναι μαθηματικά αδύνατη η αντίστροφη εξαγωγή πληροφορίας για συγκεκριμένο μεμονωμένο χρήστη, με αντάλλαγμα μια μετρήσιμη υποβάθμιση της ακρίβειας του μοντέλου: ένα ακόμη trade-off συναφές με αυτό της Ενότητας 9.11.

8.16 Πρόσθετες θεωρήσεις κανονιστικής συμμόρφωσης

Πέραν του Κανονισμού Γενικής Προστασίας Δεδομένων (GDPR) και του αναμενόμενου EU AI Act που αναφέρθηκαν στις Ενότητες 8.3 και 10.3, αναπτυσσόμενα κανονιστικά πλαίσια σε διεθνές επίπεδο επιβάλλουν αυξανόμενες απαιτήσεις διαφάνειας και λογοδοσίας για αλγοριθμικά συστήματα λήψης αποφάσεων, συμπεριλαμβανομένων των συστημάτων σύστασης σε ορισμένες κατηγορίες εφαρμογών υψηλού κινδύνου. Η ανάγκη τεκμηρίωσης της

λογικής λειτουργίας ενός αλγορίθμου σύστασης, ιδιαίτερα σε τομείς όπως η υγεία (Ενότητα 7.9) ή η εκπαίδευση (Ενότητα 7.3), ενδέχεται να απαιτήσει εκ νέου σχεδιασμό συστημάτων που βασίζονται σε δυσδιαφανείς (black-box) αρχιτεκτονικές βαθιάς μάθησης.

Η τάση αυτή ενισχύει περαιτέρω το επιχειρηματικό κίνητρο για επενδύσεις σε τεχνικές ερμηνευσιμότητας (Ενότητα 6.5) και παραγωγή εξηγήσεων (Ενότητα 8.5), καθώς η κανονιστική συμμόρφωση μετατρέπεται σταδιακά από προαιρετικό χαρακτηριστικό σε βασική προδιαγραφή σχεδίασης για συστήματα σύστασης που λειτουργούν σε ρυθμιζόμενους κλάδους, με πιθανές επιπτώσεις και στην αρχιτεκτονική επιλογή μεταξύ ακρίβειας και διαφάνειας που συζητήθηκε αναλυτικά στην Ενότητα 9.11.

8.17 Σύνοψη

Αν έπρεπε να ξεχωρίσω μία κατηγορία προκλήσεων ως πιο επείγουσα από τις υπόλοιπες, θα ήταν η ένταση ανάμεσα σε τεχνικές λύσεις και κανονιστικές απαιτήσεις: το GCL, η Federated Learning και οι αρχιτεκτονικές δύο σταδίων αντιμετωπίζουν αποτελεσματικά data sparsity, cold start και κλιμακωσιμότητα σε επίπεδο μοντέλου, αλλά η μεροληψία, η δικαιοσύνη και η ριζοσπαστικοποίηση δεν λύνονται μόνο με καλύτερο μοντέλο -- απαιτούν κανονιστικά πλαίσια που, απ' όσο είδα στη βιβλιογραφία, ακόμη προσπαθούν να προλάβουν την ταχύτητα εξέλιξης της τεχνολογίας. Η ενσωμάτωση των LLMs μάλλον θα οξύνει αυτό το χάσμα βραχυπρόθεσμα, πριν το μειώσει.

9. Συγκριτική ανάλυση μεθόδων

Αφού αναλύθηκαν εκτενώς οι κυριότερες κατηγορίες μεθόδων και τεχνικών για τα συστήματα συστάσεων στα προηγούμενα κεφάλαια, το παρόν κεφάλαιο επιχειρεί μια ολοκληρωμένη συγκριτική αποτίμησή τους. Ο στόχος είναι διπλός: αφενός να παρουσιαστούν συνοπτικά τα σημαντικότερα πειραματικά αποτελέσματα που αναφέρονται στη βιβλιογραφία, αφετέρου να αξιολογηθούν οι διαφορετικές μέθοδοι ολιστικά ως προς πολλαπλά κριτήρια: ακρίβεια, scalability, αντιμετώπιση cold start, ερμηνευσιμότητα, ποικιλομορφία και υπολογιστική πολυπλοκότητα. Η ανάλυση αυτή παρέχει ένα πρακτικό πλαίσιο επιλογής κατάλληλης μεθόδου ανάλογα με τις απαιτήσεις της εκάστοτε εφαρμογής.

9.1 Τυπικά σύνολα δεδομένων αξιολόγησης (Benchmark Datasets)

Η αξιολόγηση και σύγκριση των μεθόδων σύστασης βασίζεται σε μεγάλο βαθμό σε κοινά, δημόσια διαθέσιμα σύνολα δεδομένων αναφοράς (benchmark datasets). Η ύπαρξη κοινών datasets επιτρέπει την αντικειμενική σύγκριση διαφορετικών μεθόδων και την αξιόπιστη παρακολούθηση της προόδου του πεδίου. Ωστόσο, η σύγκριση μεθόδων σε διαφορετικά datasets είναι μη τετριμμένη, καθώς τα αποτελέσματα εξαρτώνται έντονα από τα χαρακτηριστικά του εκάστοτε dataset: πυκνότητα, μέγεθος, τομέας εφαρμογής.

Το MovieLens αποτελεί αναμφίβολα το πιο χρησιμοποιούμενο benchmark dataset στο πεδίο. Διατίθεται σε πολλαπλές εκδόσεις (100K, 1M, 10M, 25M αλληλεπιδράσεις) επιτρέποντας αξιολόγηση σε διαφορετικές κλίμακες. Τα datasets της Amazon (Electronics, Books, Movies, Beauty, κ.ά.) αντιπροσωπεύουν πραγματικά σενάρια ηλεκτρονικού εμπορίου με πλούσια μεταδεδομένα αντικειμένων, κειμενικές κριτικές και ετερογενείς τύπους αλληλεπίδρασης. Το Yelp, με κριτικές τοπικών επιχειρήσεων, και το Gowalla, ένα dataset τοποθεσίας, χρησιμοποιούνται συχνά σε GNN benchmarks. Ο Πίνακας 9.1 παρουσιάζει συνοπτικά τα κυριότερα datasets που αναφέρονται στα 100 papers που αναλύθηκαν.

Πίνακας 9.1: Κυριότερα benchmark datasets συστημάτων συστάσεων

Dataset	Χρήστες	Αντικείμενα	Αλληλεπιδράσεις	Τομέας Εφαρμογής
MovieLens 100K	943	1.682	100.000	Κινηματογράφος
MovieLens 1M	6.040	3.952	1.000.209	Κινηματογράφος
Amazon (Διάφοροι)	~100K+	~50K+	~500K+	Ηλεκτρονικό Εμπόριο
Yelp	~45K+	~20K+	~1,1M+	Τοπικές επιχειρήσεις
Retail Rocket	~1,4M	~235K	~2,7M	Ηλεκτρονικό Εμπόριο
Taobao	~48K	~39K	~2,6M	Κοινωνικό e-commerce
Food.com	~32K	~18,5K	~1,1M+	Τρόφιμα & Συνταγές
Hotel Reviews	—	—	~35K	Τουρισμός & Ξενοδοχεία

9.2 Ποσοτικά αποτελέσματα από τη βιβλιογραφία

Η συγκέντρωση και σύγκριση ποσοτικών αποτελεσμάτων από διαφορετικές μελέτες απαιτεί ιδιαίτερη προσοχή: τα αποτελέσματα εξαρτώνται από τον τύπο αξιολόγησης (offline vs. online), το πρωτόκολλο διαχωρισμού δεδομένων (train/test split ή cross-validation), τις συγκεκριμένες μετρικές που χρησιμοποιούνται και τις τιμές hyperparameters. Παρ' όλα αυτά, η παρουσίαση αντιπροσωπευτικών αποτελεσμάτων από τη βιβλιογραφία παρέχει χρήσιμη εικόνα της σχετικής απόδοσης διαφορετικών προσεγγίσεων. Ο Πίνακας 9.2 συνοψίζει τα σημαντικότερα ποσοτικά αποτελέσματα που αναφέρθηκαν στα papers που αναλύθηκαν.

Πίνακας 9.2: Συγκριτικά αποτελέσματα μοντέλων από τη βιβλιογραφία

Μοντέλο	Dataset	Μετρική	Αποτέλεσμα	Πηγή
RBM + kNN (CF)	MovieLens 1M	RMSE / MAE	0,8736 / 0,6854	Alam & Ahmed, 2025
CNN + LSTM + Attention	E-Commerce Dataset	Accuracy	89%	Yu & Aviles, 2025
Hybrid MF + DNN	RetailRocket	RMSE	Ανώτερο έναντι των CF/SVD	Wang et al., 2025
DeepMF + Sentiment	Hotel Reviews	RMSE	0,1	Darraz et al., 2025
ResNetMF (MCRS)	Τουριστικά δεδομένα	RMSE / MAE	Ανώτερο vs baselines	Payandenick et al., 2026
GCL vs GCN/GAT/LightGCN	MovieLens 100K	Precision/Diversity/Fairness	GCL ανώτερο και στα 3	Ramteke et al., 2025
HDHP (Hypergraph)	MovieLens / Amazon	Recall@20 / NDCG@20	88,76% / 42,01%	Mall & Singh, 2026
FRMADHG (Food)	Food.com / Allrecipes	Precision@10 / Recall@10	+19,8% / +18,7% vs CF	Alkhrijah et al., 2026
CBO-FFNN (Education)	Library Dataset	Accuracy / F-value	91,4% / 95,7%	Bian & Chang, 2026
MycGNN	RetailRocket / ML-1M	Accuracy / Diversity	Υψηλή diversity vs SOTA	Bahi et al., 2026

Μια σημαντική παρατήρηση από τα δεδομένα του Πίνακα 9.2 είναι ότι οι μέθοδοι βαθιάς μάθησης και γραφικών νευρωνικών δικτύων υπερτερούν σταθερά των κλασικών μεθόδων CF και Matrix Factorization ως προς ακρίβεια. Αξιοσημείωτο είναι επίσης ότι τα υβριδικά μοντέλα (που συνδυάζουν δύο ή περισσότερες τεχνικές) τείνουν να αποδίδουν καλύτερα από κάθε μέθοδο μόνη της. Το μοντέλο DeepMF με ενσωμάτωση sentiment analysis πέτυχε εξαιρετικά χαμηλό RMSE (0,1), καταδεικνύοντας την αξία της αξιοποίησης επιπλέον πηγών πληροφορίας. Το HDHP hypergraph μοντέλο πέτυχε Recall@20 = 88,76% (εξαιρετικά υψηλή τιμή για βάση κοινωνικής σύστασης) με παράλληλη βελτίωση 83,95% σε σχέση με το baseline DHCF.

9.3 Πολυδιάστατη συγκριτική αξιολόγηση

Η μονοδιάστατη σύγκριση μεθόδων αποκλειστικά βάσει ακρίβειας είναι ανεπαρκής για πρακτικές εφαρμογές. Ένα σύστημα με εξαιρετική ακρίβεια αλλά μηδενική ερμηνευσιμότητα δεν είναι κατάλληλο για εφαρμογές υψηλής ευθύνης όπως η υγεία. Ένα σύστημα με υψηλή ακρίβεια αλλά χαμηλή scalability δεν μπορεί να λειτουργήσει σε πλατφόρμα εκατομμυρίων χρηστών. Για τον λόγο αυτό, ο Πίνακας 9.3 παρέχει πολυδιάστατη αξιολόγηση των κυριότερων κατηγοριών μεθόδων.

Πίνακας 9.3: Πολυδιάστατη συγκριτική αξιολόγηση κατηγοριών μεθόδων

Κατηγορία	Ακρίβεια	Κλιμακωσιμότητα	Cold Start	Ερμηνευσιμότητα	Ποικιλομορφία	Πολυπλοκότητα
CF (Memory-based)	Μέτρια	Χαμηλή	Κακή	Υψηλή	Χαμηλή	Χαμηλή
Matrix Factorization	Καλή	Μέτρια	Μέτρια	Μέτρια	Χαμηλή	Χαμηλή
NCF / MLP	Καλή	Μέτρια	Μέτρια	Χαμηλή	Χαμηλή	Μέτρια
RNN / LSTM	Καλή	Μέτρια	Μέτρια	Χαμηλή	Μέτρια	Μέτρια
Transformer	Πολύ Καλή	Χαμηλή	Καλή	Μέτρια	Μέτρια	Υψηλή
GNN (LightGCN κ.ά.)	Πολύ Καλή	Καλή	Μέτρια	Χαμηλή	Μέτρια	Μέτρια
GCL / SSL	Πολύ Καλή	Καλή	Καλή	Χαμηλή	Καλή	Μέτρια
Hybrid (DL+CF)	Πολύ Καλή	Μέτρια	Καλή	Μέτρια	Μέτρια	Υψηλή
LLM-based	Πολύ Καλή	Πολύ Χαμηλή	Πολύ Καλή	Πολύ Υψηλή	Καλή	Πολύ Υψηλή

9.4 Ανάλυση ανά διάσταση αξιολόγησης

9.4.1 Ακρίβεια προβλέψεων

Ως προς την ακρίβεια προβλέψεων, η εξέλιξη είναι σαφής: κάθε γενιά μεθόδων βελτιώνει σταθερά τις προηγούμενες. Η κλασική memory-based CF αποτελεί το χαμηλότερο σημείο αναφοράς. Η Matrix Factorization σηματοδότησε το πρώτο μεγάλο άλμα. Τα μοντέλα βαθιάς μάθησης (NCF, RNN, CNN, Transformer) βελτίωσαν περαιτέρω την ακρίβεια αξιοποιώντας μη γραμμικές αλληλεπιδράσεις. Τα GNN-based μοντέλα αξιοποιούν τη δομή του γράφου και πετυχαίνουν state-of-the-art αποτελέσματα σε μεγάλα benchmarks. Το Graph Contrastive Learning (GCL) δείχνει ανώτερη συνολική απόδοση όταν συνυπολογίζεται και η αντιμετώπιση προβλημάτων cold start και long-tail, σε αντίθεση με τα κλασικά GNN που υπερτερούν μόνο σε πυκνά δεδομένα.

9.4.2 Κλιμακωσιμότητα

Ως προς την scalability, τα memory-based μοντέλα (user-based CF, item-based CF) είναι τα πιο αδύναμα, με πολυπλοκότητα $O(|U|^2)$ ή $O(|I|^2)$ αντίστοιχα. Τα μοντέλα MF και NCF εκπαιδεύονται αποτελεσματικά σε μεγάλα δεδομένα, αλλά η πρόβλεψη σε πραγματικό χρόνο απαιτεί επιπλέον βελτιστοποίηση. Τα GNN και ιδιαίτερα το LightGCN (με την απλοποιημένη γραφική διάδοση) ισορροπούν ικανοποιητικά ακρίβεια και κλιμακωσιμότητα. Τα Transformer-based μοντέλα έχουν θεωρητική πολυπλοκότητα $O(n^2)$ ως προς το μήκος ακολουθίας, κάτι που περιορίζει την εφαρμογή τους σε πολύ μεγάλες ακολουθίες χωρίς προσεγγιστικές τεχνικές attention. Τα LLM-based μοντέλα έχουν τη χαμηλότερη κλιμακωσιμότητα λόγω του τεράστιου αριθμού παραμέτρων τους.

9.4.3 Αντιμετώπιση cold start

Ως προς το cold start, τα κλασικά CF μοντέλα αντιμετωπίζουν πλήρη αποτυχία: χωρίς ιστορικό αλληλεπίδρασης, το σύστημα δεν μπορεί να παράγει συστάσεις. Τα content-based μοντέλα αντιμετωπίζουν μερικά το cold start νέων χρηστών αλλά όχι νέων αντικειμένων χωρίς περιγραφή. Τα Knowledge Graph-enhanced μοντέλα αντιμετωπίζουν το cold start αντικειμένων αξιοποιώντας σημασιολογικές πληροφορίες. Τα GCL/SSL μοντέλα βελτιώνουν σημαντικά την απόδοση για long-tail χρήστες και αντικείμενα. Τα LLM-based μοντέλα έχουν ίσως τη μεγαλύτερη δυνατότητα zero-shot γενίκευσης, αξιοποιώντας ευρεία πρότερη γνώση.

9.4.4 Ερμηνευσιμότητα

Ως προς την ερμηνευσιμότητα, υπάρχει σαφής αντίστροφη σχέση με την ακρίβεια: τα πιο ακριβή μοντέλα τείνουν να είναι τα λιγότερο ερμηνεύσιμα. Η κλασική CF και η content-based filtering παρέχουν φυσικές εξηγήσεις («σας αρέσουν ταινίες του ίδιου σκηνοθέτη»), ενώ τα βαθιά νευρωνικά δίκτυα λειτουργούν ως black boxes. Τα GAT και άλλα attention-based μοντέλα παρέχουν ημι-ερμηνεύσιμα αποτελέσματα μέσω των βαρών attention. Τα LLM-based μοντέλα όπως το STLLM-Rec ανοίγουν νέες δυνατότητες ερμηνευσιμότητας μέσω φυσικής γλώσσας, αλλά η αξιοπιστία αυτών των εξηγήσεων παραμένει ανοιχτό ζήτημα.

9.4.5 Ποικιλομορφία

Ως προς την ποικιλομορφία (diversity) των συστάσεων, η πλειοψηφία των μεθόδων που βελτιστοποιούν ακρίβεια τείνουν να θυσιάζουν ποικιλομορφία. Το MycGNN αποτελεί εξαίρεση, εισάγοντας ρητά έναν μηχανισμό exploration-exploitation που ισορροπεί ποικιλομορφία και ακρίβεια. Το GCL αντιμετωπίζει εν μέρει τη μεροληψία δημοτικότητας (popularity bias) και βελτιώνει την ποικιλομορφία για long-tail αντικείμενα. Γενικότερα, η ποικιλομορφία παραμένει ένα πεδίο που χρειάζεται ρητή βελτιστοποίηση: η βελτίωσή της δεν προκύπτει αυτόματα από τη βελτίωση της ακρίβειας.

9.5 Πρακτικές κατευθυντήριες γραμμές επιλογής μεθόδου

Βάσει της συγκριτικής ανάλυσης, προκύπτουν ορισμένες πρακτικές κατευθύνσεις για την επιλογή κατάλληλης μεθόδου ανάλογα με το σενάριο εφαρμογής. Για πλατφόρμες πολύ μεγάλης κλίμακας (εκατομμύρια χρήστες/αντικείμενα) όπου η ταχύτητα είναι κρίσιμη, τα LightGCN-τύπου μοντέλα με αρχιτεκτονική δύο σταδίων (retrieval + ranking) αποτελούν ισορροπημένη επιλογή. Για περιβάλλοντα με σοβαρό cold start πρόβλημα, τα Knowledge Graph-enhanced μοντέλα ή τα LLM-based προσφέρουν καλύτερη γενίκευση. Για εφαρμογές που απαιτούν ερμηνευσιμότητα (π.χ. υγεία, χρηματοοικονομικά) τα attention-based ή LLM-based μοντέλα με ρητή εξήγηση είναι προτιμότερα.

Για εφαρμογές με πλούσια δεδομένα κοινωνικού δικτύου, τα trust-based και heterogeneous graph μοντέλα αξιοποιούν αποτελεσματικά τις κοινωνικές πληροφορίες. Για σενάρια sequential recommendation (e-commerce, streaming), τα SASRec/BERT4Rec Transformer-based μοντέλα παρέχουν υψηλή ακρίβεια στη σύλληψη βραχυπρόθεσμων και μακροπρόθεσμων ενδιαφερόντων. Τέλος, για εφαρμογές όπου η ποικιλομορφία και η αντιμετώπιση φυσαλίδων φίλτρων είναι προτεραιότητα (π.χ. ειδησεογραφία, εκπαίδευση, κοινωνικά δίκτυα) τα diversity-aware μοντέλα όπως το MycGNN ή τα fairness-aware μοντέλα όπως το FairDA αξίζουν ιδιαίτερης προσοχής. Σε κάθε περίπτωση, τα υβριδικά σχήματα που συνδυάζουν παραπάνω από μία προσεγγίσεις τείνουν να αποδίδουν ανώτερα από κάθε μέθοδο μεμονωμένα.

9.6 Ανάλυση cross-domain και ειδικών τομέων εφαρμογής

Η συγκριτική ανάλυση του Κεφαλαίου 9 θα ήταν ελλιπής χωρίς αναφορά στην απόδοση μοντέλων σε ειδικές κατηγορίες εφαρμογών που δεν αντιπροσωπεύονται επαρκώς από τα κλασικά benchmark datasets. Τα cross-domain μοντέλα, τα ψυχολογικά μοντέλα και τα εξειδικευμένα domain-specific συστήματα παρουσιάζουν ιδιαίτερες δυνατότητες που δεν αναδεικνύονται σε γενικά benchmarks όπως το MovieLens ή το Gowalla.

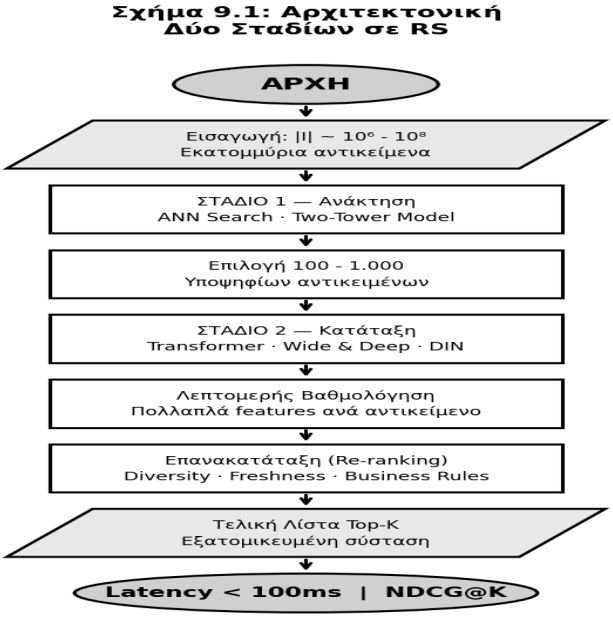
Συγκεκριμένα, τα cross-domain μοντέλα είναι σχεδιασμένα για να αξιοποιούν την αραιότητα: επιδεικνύουν μεγαλύτερα κέρδη ακριβώς εκεί όπου τα single-domain μοντέλα αποτυγχάνουν: σε νέους χρήστες και αντικείμενα χωρίς ιστορικό. Τα ψυχολογικά μοντέλα όπως το NeuroGraph-CPM εξελίσσονται σε κατεύθυνση υψηλότερης personalization, πέρα από

επιφανειακή συμπεριφορική ανάλυση. Τα domain-specific συστήματα (CARES, CO_2 RS, Manufacturing RS) αναδεικνύουν ότι η εξειδίκευση για συγκεκριμένα πεδία μπορεί να αποφέρει ιδιαίτερα σημαντικά αποτελέσματα που δεν είναι συγκρίσιμα με γενικά μετρικά ανάλυσης.

Συνολικά, η εξέταση των ειδικών τομέων καταδεικνύει ότι δεν υπάρχει «ένα μοντέλο για όλες τις περιπτώσεις»: η βέλτιστη αρχιτεκτονική εξαρτάται ουσιαστικά από τα χαρακτηριστικά δεδομένων, τον τομέα εφαρμογής, τις απαιτήσεις ερμηνευσιμότητας και τους περιορισμούς κλιμακωσιμότητας (scalability). Η επιλογή μεθόδου πρέπει να αντιμετωπίζεται ως πολυκριτηριακή απόφαση, όπου η ακρίβεια είναι μόνο ένα από τα πολλά κριτήρια.

9.7 Αρχιτεκτονική δύο σταδίων σε πλατφόρμες μεγάλης κλίμακας

Το Σχήμα 9.1 παρουσιάζει την αρχιτεκτονική δύο σταδίων που χρησιμοποιείται από μεγάλες πλατφόρμες (YouTube, Amazon, TikTok) για να επιτύχουν ταυτόχρονα υψηλή ακρίβεια και latency < 100ms. Στο πρώτο στάδιο (Retrieval), αλγόριθμοι Approximate Nearest Neighbor επιλέγουν γρήγορα 100-1000 υποψήφια αντικείμενα από εκατομμύρια. Στο δεύτερο στάδιο (Ranking), ένα βαρύτερο DNN μοντέλο αξιολογεί λεπτομερώς τους υποψηφίους. Ακολουθεί re-ranking για diversity και freshness.



Σχήμα 9.1: Αρχιτεκτονική δύο σταδίων (Retrieval + Ranking) για κλιμακώσιμα RS

9.8 Στατιστική σημαντικότητα στις συγκρίσεις μοντέλων

Οι συγκριτικές αξιολογήσεις που παρουσιάστηκαν στις προηγούμενες ενότητες του παρόντος κεφαλαίου βασίζονται σε μεμονωμένες μετρήσεις απόδοσης (π.χ. NDCG@10, Recall@20) που αναφέρονται στη βιβλιογραφία. Μια σημαντική μεθοδολογική παράμετρος που συχνά παραλείπεται είναι η στατιστική σημαντικότητα των παρατηρούμενων διαφορών: μια διαφορά της τάξης του 1-2% σε ένα benchmark dataset δεν είναι απαραίτητα στατιστικά σημαντική, ιδιαίτερα όταν τα πειράματα δεν περιλαμβάνουν πολλαπλές επαναλήψεις με διαφορετικούς αρχικοποιητές τυχαιότητας (random seeds).

Στη βιβλιογραφία μηχανικής μάθησης προτείνεται η χρήση στατιστικών ελέγχων όπως ο ζευγαρωτός t-test ή, προτιμότερα λόγω της μη παραμετρικής φύσης των κατανομών μετρικών κατάταξης, ο έλεγχος Wilcoxon signed-rank, για τη σύγκριση δύο μοντέλων στο ίδιο σύνολο δοκιμής. Για συγκρίσεις πολλαπλών μοντέλων ταυτόχρονα, απαιτείται διόρθωση για πολλαπλές συγκρίσεις (π.χ. διόρθωση Bonferroni), ώστε να αποφευχθεί η αυξημένη πιθανότητα ψευδώς

θετικών ευρημάτων (Type I error) λόγω του μεγάλου αριθμού ζευγαρωτών συγκρίσεων.

Η απουσία συστηματικής αναφοράς στατιστικής σημαντικότητας αποτελεί γνωστό μεθοδολογικό κενό σε μεγάλο μέρος της βιβλιογραφίας συστημάτων συστάσεων, όπως αναδεικνύεται και από συστηματικές ανασκοπήσεις του πεδίου. Για τον λόγο αυτό, οι συγκρίσεις των πινάκων του παρόντος κεφαλαίου και του Παραρτήματος Β θα πρέπει να ερμηνεύονται ως ενδεικτικές τάσεις απόδοσης και όχι ως αυστηρά στατιστικά τεκμηριωμένα συμπεράσματα, μια επιφύλαξη που ισχύει ευρέως στη σύγχρονη βιβλιογραφία του πεδίου.

9.9 Η μεθοδολογική σημασία των ablation studies

Πέραν της απευθείας σύγκρισης διαφορετικών μοντέλων, τα ablation studies (μελέτες αφαίρεσης συστατικών) αποτελούν θεμελιώδες εργαλείο για την κατανόηση της συμβολής κάθε επιμέρους συστατικού μιας σύνθετης αρχιτεκτονικής στη συνολική απόδοση. Για παράδειγμα, σε ένα μοντέλο τύπου NeuMF (Ενότητα 4.1), ένα ablation study θα συνέκρινε την πλήρη αρχιτεκτονική με εκδοχές που διαθέτουν μόνο τον κλάδο GMF, μόνο τον κλάδο MLP, ή διαφορετικό αριθμό επιπέδων, ώστε να τεκμηριωθεί ποσοτικά η αξία της αρχιτεκτονικής σύντηξης (fusion) των δύο κλάδων.

Στο πλαίσιο των GNN αρχιτεκτονικών, τα ablation studies διερευνούν συνήθως την επίδραση του αριθμού επιπέδων διάδοσης (συνδέοντάς το άμεσα με το φαινόμενο over-smoothing της Ενότητας 5.12), της παρουσίας ή απουσίας κανονικοποίησης (regularization), και της επιλογής συνάρτησης απώλειας (π.χ. BPR έναντι binary cross-entropy). Η απουσία συστηματικών ablation studies σε μια δημοσίευση αποτελεί συχνά ένδειξη ότι η αναφερόμενη βελτίωση απόδοσης μπορεί να αποδίδεται σε δευτερεύοντες παράγοντες (π.χ. καλύτερο tuning υπερπαραμέτρων) και όχι στην πραγματική αρχιτεκτονική καινοτομία που προτείνεται, ζήτημα που σχετίζεται άμεσα με την κρίση αναπαραγωγιμότητας της Ενότητας 8.13.

9.10 Trade-off ακρίβειας και ερμηνευσιμότητας

Ο Πίνακας 9.4 συνοψίζει μια γενική τάση που παρατηρείται διαχρονικά στη βιβλιογραφία: οι πιο ακριβείς μέθοδοι σύστασης τείνουν να είναι και οι λιγότερο ερμηνεύσιμες, δημιουργώντας ένα βασικό trade-off που πρέπει να σταθμίσει κάθε σχεδιαστής συστήματος ανάλογα με τις απαιτήσεις της εφαρμογής: ιδιαίτερα κρίσιμο σε τομείς υψηλού κινδύνου όπως η υγεία (Ενότητα 7.9), όπου η ερμηνευσιμότητα δεν είναι απλώς επιθυμητή αλλά συχνά κανονιστικά απαραίτητη.

Πίνακας 9.4: Ενδεικτικό trade-off ακρίβειας, ερμηνευσιμότητας και υπολογιστικού κόστους

Μέθοδος	Ακρίβεια	Ερμηνευσιμότητα	Υπολ. Κόστος
User/Item-based CF	Μέτρια	Υψηλή	Χαμηλό
Matrix Factorization	Καλή	Μέτρια	Χαμηλό
NCF / NeuMF	Πολύ καλή	Χαμηλή	Μέτριο
LightGCN	Πολύ καλή - Άριστη	Χαμηλή	Μέτριο-Υψηλό
SASRec (Transformer)	Άριστη	Πολύ χαμηλή	Υψηλό

9.11 Συνολική σύνθεση: πότε επιλέγεται κάθε οικογένεια μεθόδων

Συνθέτοντας τις συγκριτικές αναλύσεις των προηγούμενων ενοτήτων, μπορεί να διατυπωθεί ένας γενικός πρακτικός οδηγός επιλογής αρχιτεκτονικής. Για εφαρμογές με περιορισμένους υπολογιστικούς πόρους, μικρό όγκο δεδομένων ή ανάγκη για υψηλή ερμηνευσιμότητα (π.χ. ρυθμιζόμενοι τομείς όπως η χρηματοοικονομία ή η υγεία, Ενότητα 7.9), οι κλασικές μέθοδοι συνεργατικής διήθησης ή παραγοντοποίησης πίνακα (Κεφάλαιο 3) παραμένουν συχνά η πιο κατάλληλη επιλογή, παρά τη χαμηλότερη θεωρητική ακρίβειά τους.

Για εφαρμογές με μεγάλο όγκο δεδομένων αλληλεπίδρασης και σαφή σειριακή δομή συμπεριφοράς (π.χ. ηλεκτρονικό εμπόριο, ροή περιεχομένου), οι αρχιτεκτονικές Transformer-based, όπως το SASRec, προσφέρουν συνήθως την καλύτερη ακρίβεια, με αντάλλαγμα το αυξημένο υπολογιστικό κόστος. Οι αρχιτεκτονικές GNN-based, όπως το LightGCN, αποτελούν συχνά την ισορροπημένη ενδιάμεση επιλογή, ιδιαίτερα όταν υπάρχει πλούσια δομή γράφου (π.χ. κοινωνικές σχέσεις, Ενότητα 4.16) που μπορεί να αξιοποιηθεί. Η τελική επιλογή πρέπει

πάντα να σταθμίζει τους παράγοντες ακρίβειας, ερμηνευσιμότητας, υπολογιστικού κόστους και διαθέσιμων δεδομένων που αναλύθηκαν στο παρόν κεφάλαιο, χωρίς να υπάρχει μία «καθολικά βέλτιστη» αρχιτεκτονική για όλα τα σενάρια εφαρμογής.

9.12 Όρια γενίκευσης των συγκριτικών ευρημάτων

Τα συγκριτικά ευρήματα που παρουσιάστηκαν στο παρόν κεφάλαιο βασίζονται σε αναφερόμενα αποτελέσματα πάνω σε συγκεκριμένα, ευρέως χρησιμοποιούμενα benchmark datasets (Πίνακας Β.5), τα οποία, ωστόσο, δεν είναι απαραίτητα αντιπροσωπευτικά όλων των πιθανών πραγματικών σεναρίων εφαρμογής. Datasets όπως το MovieLens ή το Amazon Reviews χαρακτηρίζονται από συγκεκριμένα στατιστικά χαρακτηριστικά (πυκνότητα, κατανομή δημοτικότητας, μέγεθος καταλόγου) που ενδέχεται να διαφέρουν σημαντικά από τα χαρακτηριστικά μιας συγκεκριμένης πραγματικής εφαρμογής, περιορίζοντας την άμεση μεταφερσιμότητα των συγκριτικών ευρημάτων.

Επιπλέον, η σχετική κατάταξη μεθόδων μπορεί να μεταβάλλεται σημαντικά ανάλογα με τη μετρική αξιολόγησης που επιλέγεται (Ενότητα 2.12): ένα μοντέλο που υπερτερεί ως προς το NDCG@10 μπορεί να υπολείπεται ως προς τη διαφοροποίηση (diversity) ή την κάλυψη καταλόγου. Για τον λόγο αυτό, η επιλογή αρχιτεκτονικής για μια συγκεκριμένη πραγματική εφαρμογή θα πρέπει πάντα να συνοδεύεται από στοχευμένη αξιολόγηση στα πραγματικά δεδομένα της εφαρμογής, χρησιμοποιώντας τα ευρήματα της βιβλιογραφίας ως αρχικό, ενδεικτικό σημείο αναφοράς και όχι ως οριστικό κριτήριο επιλογής.

9.13 Συνοπτικός πίνακας τελικών συστάσεων επιλογής

Με βάση το σύνολο της συγκριτικής ανάλυσης του παρόντος κεφαλαίου, ο Πίνακας 9.5 παρέχει μια τελική, συνοπτική σύσταση επιλογής αρχιτεκτονικής ανά σενάριο εφαρμογής, συνθέτοντας τα κριτήρια ακρίβειας, ερμηνευσιμότητας, υπολογιστικού κόστους και διαθέσιμων δεδομένων που αναλύθηκαν στις Ενότητες 9.1 έως 9.12.

Πίνακας 9.5: Συνοπτικός οδηγός επιλογής αρχιτεκτονικής ανά σενάριο εφαρμογής

Σενάριο Εφαρμογής	Προτεινόμενη Οικογένεια Μεθόδων
Περιορισμένα δεδομένα / πόροι	Κλασική CF ή Matrix Factorization (Κεφάλαιο 3)
Ανάγκη υψηλής ερμηνευσιμότητας	User/Item-based CF ή υβριδικά γραμμικά μοντέλα (Ενότητα 3.13)
Μεγάλος όγκος σειριακών δεδομένων	SASRec / BERT4Rec (Ενότητες 4.5.1, 4.14)
Πλούσια δομή γράφου / κοινωνικά δεδομένα	LightGCN ή ετερογενή GNN (Ενότητες 5.2, 5.4, 5.13)
Ανάγκη ανθεκτικότητας σε αραιότητα	VAE-based ή GCL-based μοντέλα (Ενότητες 4.13, 4.7, 5.6, 5.11)
Πολλαπλά ενδιαφερόμενα μέρη	Re-ranking με περιορισμούς δικαιοσύνης (Ενότητα 6.10)

9.14 Σύνοψη

Η ποσοτική σύγκριση επιβεβαίωσε το αναμενόμενο -- ότι τα μοντέλα βαθιάς μάθησης και GNN υπερτερούν σταθερά των κλασικών μεθόδων ως προς ακρίβεια, με τα υβριδικά σχήματα στην κορυφή -- αλλά αυτό που θεωρώ πιο χρήσιμο εύρημα του κεφαλαίου είναι ότι καμία μέθοδος δεν αριστεύει σε όλες τις διαστάσεις ταυτόχρονα (ακρίβεια, scalability, αντιμετώπιση cold start, ερμηνευσιμότητα, ποικιλομορφία). Αυτό σημαίνει πρακτικά ότι η ερώτηση «ποιο μοντέλο είναι το καλύτερο» δεν έχει νόημα χωρίς να προσδιοριστεί πρώτα ποια διάσταση προτεραιοποιεί η συγκεκριμένη εφαρμογή· τα κριτήρια επιλογής που παρουσιάστηκαν έχουν αξία ακριβώς επειδή αναγκάζουν σε αυτή την προτεραιοποίηση.

10. Συμπεράσματα

Η παρούσα πτυχιακή εργασία είχε ως αντικείμενο τη συστηματική αποτύπωση και κριτική ανάλυση των ευφυών τεχνικών που χρησιμοποιούνται για την ανάπτυξη σύγχρονων συστημάτων συστάσεων. Μέσα από την ανάλυση εκατό επιστημονικών άρθρων δημοσιευμένων κατά κύριο λόγο από το 2020 έως το 2026, χαρτογραφήθηκε η εξέλιξη του πεδίου από τις κλασικές μεθόδους έως τις πιο σύγχρονες αρχιτεκτονικές, αναδείχθηκαν τα δυνατά και αδύναμα σημεία κάθε προσέγγισης, και αναλύθηκαν οι ανοιχτές ερευνητικές προκλήσεις. Το παρόν κεφάλαιο συνοψίζει τα κεντρικά ευρήματα, αναδεικνύει τη συνεισφορά της εργασίας και υποδεικνύει κατευθύνσεις για μελλοντική έρευνα.

10.1 Κεντρικά ευρήματα

Το πρώτο κεντρικό εύρημα της εργασίας είναι ότι η εξέλιξη των συστημάτων συστάσεων ακολουθεί μια σαφή τροχιά αύξουσας πολυπλοκότητας και ακρίβειας, που παράλληλα εγείρει ολοένα σημαντικότερες προκλήσεις κλιμακωσιμότητας (scalability) και ερμηνευσιμότητας. Οι κλασικές μέθοδοι (collaborative filtering και matrix factorization) διαμόρφωσαν τα θεμέλια του πεδίου και παραμένουν σημεία αναφοράς, αλλά η αδυναμία τους να συλλάβουν μη γραμμικές σχέσεις τις καθιστά ανεπαρκείς για τις απαιτήσεις των σύγχρονων εφαρμογών.

Το δεύτερο εύρημα αφορά τον μετασχηματιστικό ρόλο της βαθιάς μάθησης. Αρχιτεκτονικές όπως τα NCF, RNN/LSTM, CNN και Transformers ξεπέρασαν τους περιορισμούς των γραμμικών μεθόδων, μαθαίνοντας αυτόματα πολύπλοκες, ιεραρχικές αναπαραστάσεις. Ιδιαίτερα τα Transformers με τον μηχανισμό self-attention επέλυσαν το πρόβλημα των μακροπρόθεσμων εξαρτήσεων που ταλάνιζε τα RNN, καθιστώντας μοντέλα όπως το SASRec και το BERT4Rec state-of-the-art για σειριακή σύσταση. Παράλληλα, τα Γραφικά Νευρωνικά Δίκτυα εισήγαγαν μια θεμελιωδώς νέα οπτική: η μοντελοποίηση των αλληλεπιδράσεων ως γράφος επιτρέπει τη σύλληψη υψηλής τάξης σχέσεων που οι παραδοσιακές μέθοδοι αδυνατούν να εντοπίσουν.

Τρίτο σημαντικό εύρημα είναι η αναδυόμενη σημασία της αυτο-επιβλεπόμενης μάθησης (SSL) και της αντιπαραθετικής μάθησης (Contrastive Learning) ως εργαλείων αντιμετώπισης της αραιότητας δεδομένων και της μεροληψίας δημοτικότητας. Το Graph Contrastive Learning αποδείχθηκε ανώτερο από τα κλασικά GNN μοντέλα όχι μόνο σε ακρίβεια αλλά και σε ποικιλομορφία και δικαιοσύνη (τρεις διαστάσεις κρίσιμες για εφαρμογές πραγματικού κόσμου) χωρίς να απαιτεί επιπλέον ετικετιαρισμένα δεδομένα.

Τέταρτο εύρημα αφορά την πολυδιάστατη φύση των σύγχρονων συστημάτων συστάσεων. Η ανάλυση ανέδειξε ότι η ακρίβεια είναι μόνο ένα από τα κριτήρια αξιολόγησης: ποικιλομορφία, ερμηνευσιμότητα, δικαιοσύνη, ιδιωτικότητα και ανθεκτικότητα σε επιθέσεις αποτελούν εξίσου κρίσιμες διαστάσεις. Τα μοντέλα που αγνοούν αυτές τις διαστάσεις (ακόμα και αν πετυχαίνουν εξαιρετική ακρίβεια σε benchmark datasets) μπορεί να προκαλέσουν σοβαρή βλάβη όταν εφαρμοστούν σε πραγματικά περιβάλλοντα.

Πέμπτο και τελευταίο κεντρικό εύρημα είναι η διαμορφούμενη νέα τάξη πραγμάτων με την είσοδο των Μεγάλων Γλωσσικών Μοντέλων (LLMs) στο πεδίο. Μοντέλα όπως το STLLM-Rec δείχνουν ότι η ολοκλήρωση LLMs μπορεί να αντιμετωπίσει ταυτόχρονα ζητήματα ερμηνευσιμότητας, cold start και zero-shot γενίκευσης. Ωστόσο, το υψηλό υπολογιστικό κόστος, η τάση για ψευδοεξηγήσεις (hallucination) και οι δυσκολίες ευθυγράμμισης με τα δεδομένα σύστασης καθιστούν αυτό το πεδίο ακόμα ανοιχτό και σε ενεργή εξέλιξη.

Επιπρόσθετα, η συστηματική κατηγοριοποίηση των εκατό άρθρων ανά αρχιτεκτονική, τομέα εφαρμογής και έτος δημοσίευσης ανέδειξε σαφείς τάσεις μετατόπισης του ερευνητικού ενδιαφέροντος: από τις κλασικές μεθόδους συνεργατικής διήθησης και παραγοντοποίησης πίνακα προς αρχιτεκτονικές βαθιάς μάθησης, και πιο πρόσφατα προς GNN-based και Transformer-based προσεγγίσεις, με παράλληλη αύξηση του ενδιαφέροντος για ζητήματα δικαιοσύνης, ερμηνευσιμότητας και υπεύθυνης τεχνητής νοημοσύνης στο πεδίο, όπως αναλύθηκε εκτενώς στο Κεφάλαιο 8.

10.2 Συνεισφορά της εργασίας

Η παρούσα εργασία συνεισφέρει στη βιβλιογραφία με τρεις κύριους τρόπους. Πρώτον, παρέχει μια εκτεταμένη και ενημερωμένη βιβλιογραφική επισκόπηση του πεδίου, καλύπτοντας εκατό επιστημονικά άρθρα από διαφορετικές κατηγορίες μεθόδων και τομείς εφαρμογής. Αυτή η ευρύτητα κάλυψης παρέχει στον αναγνώστη μια ολοκληρωμένη εικόνα της σύγχρονης κατάστασης του πεδίου, η οποία δεν εντοπίζεται εύκολα σε μεμονωμένες επισκοπήσεις που εστιάζουν σε μία μόνο κατηγορία μεθόδων.

Δεύτερον, η εργασία παρέχει μια δομημένη, πολυδιάστατη συγκριτική ανάλυση που αξιολογεί τις μεθόδους ταυτόχρονα ως προς ακρίβεια, scalability, αντιμετώπιση cold start, ερμηνευσιμότητα, ποικιλομορφία και υπολογιστική πολυπλοκότητα. Αυτό το πλαίσιο αξιολόγησης παρέχει πρακτική καθοδήγηση για μηχανικούς και ερευνητές που καλούνται να επιλέξουν κατάλληλη μέθοδο για συγκεκριμένη εφαρμογή.

Τρίτον, η εργασία αναδεικνύει και αναλύει τις ηθικές, κοινωνικές και νομικές διαστάσεις των συστημάτων συστάσεων (ιδιωτικότητα, μεροληψία, ριζοσπαστικοποίηση, ασφάλεια) σε ισότιμη βάση με τις τεχνικές. Αυτή η ισορροπημένη προσέγγιση αντικατοπτρίζει τη σύγχρονη κατανόηση ότι η ανάπτυξη υπεύθυνης τεχνητής νοημοσύνης απαιτεί να συνυπολογίζονται οι κοινωνικές επιπτώσεις εξίσου με την τεχνική απόδοση.

10.3 Κατευθύνσεις για μελλοντική έρευνα

Βάσει της ανάλυσης που διεξήχθη, αναδεικνύονται αρκετές ερευνητικές κατευθύνσεις που αξίζουν ιδιαίτερης προσοχής στα επόμενα χρόνια. Η πρώτη και ίσως πιο επείγουσα αφορά την ανάπτυξη αποτελεσματικών μεθόδων ολοκλήρωσης LLMs με συστήματα συστάσεων, που να αντιμετωπίζουν ταυτόχρονα το υψηλό υπολογιστικό κόστος, το φαινόμενο hallucination και τις δυσκολίες ευθυγράμμισης. Η self-training προσέγγιση του STLLM-Rec αποτελεί ενθαρρυντικό πρώτο βήμα, αλλά χρειάζεται περαιτέρω αξιολόγηση σε ευρεία κλίμακα.

Δεύτερη κατεύθυνση αφορά την ανάπτυξη συστημάτων συστάσεων που ενσωματώνουν εγγενώς τις αρχές ιδιωτικότητας (privacy-by-design) αντί να τις προσθέτουν ως μεταγενέστερη επέκταση. Η Federated Learning και η Differential Privacy αποτελούν υποσχόμενα εργαλεία, ωστόσο η πρακτική εφαρμογή τους σε σύνθετες αρχιτεκτονικές GNN και Transformer δεν φαίνεται να έχει ξεπεράσει ακόμη το στάδιο των εργαστηριακών πειραμάτων στην εξεταζόμενη βιβλιογραφία. Τρίτη κατεύθυνση αφορά την ανάπτυξη ενοποιημένων πλαισίων που αντιμετωπίζουν ταυτόχρονα ακρίβεια, δικαιοσύνη και ποικιλομορφία: αντί να τις αντιμετωπίζουν ως ανταγωνιστικούς στόχους, να τις αντιμετωπίζουν ως συμπληρωματικές.

Τέταρτη κατεύθυνση αφορά την ανάπτυξη πιο αξιόπιστων μεθόδων αξιολόγησης, πέρα από τα offline benchmarks. Τα offline πειράματα σε datasets όπως το MovieLens δεν συλλαμβάνουν σημαντικές πτυχές της απόδοσης πραγματικού κόσμου: π.χ. την επίδραση των συστάσεων στη μακροπρόθεσμη συμπεριφορά χρηστών, τις κοινωνικές επιπτώσεις της φυσαλίδας φίλτρων ή την αλληλεπίδραση μεταξύ συστάσεων και αγοραστικής συμπεριφοράς. Online πειράματα (A/B testing) και μέθοδοι αξιολόγησης πέρα από την ακρίβεια αποτελούν αναγκαία εξέλιξη. Τέλος, πέμπτη κατεύθυνση αφορά τα GNN-Transformer υβριδικά μοντέλα: η ενσωμάτωση της δυνατότητας σύλληψης παγκόσμιων εξαρτήσεων των Transformers με τη γραφική αναπαράσταση των GNN υπόσχεται ανώτερη απόδοση από κάθε αρχιτεκτονική ξεχωριστά, χωρίς τα μειονεκτήματα του over-smoothing.

Μια ιδιαίτερα ενεργή κατεύθυνση έρευνας αφορά την ενσωμάτωση Μεγάλων Γλωσσικών Μοντέλων (LLMs) στον πυρήνα των συστημάτων συστάσεων, είτε ως μηχανισμό παραγωγής εξηγήσεων σε φυσική γλώσσα, είτε ως θεμελιωτικό μοντέλο (foundation model) ικανό να γενικεύει σε πολλαπλούς τομείς εφαρμογής χωρίς εκπαίδευση από το μηδέν για κάθε νέο σύνολο δεδομένων. Η προσέγγιση αυτή θα μπορούσε δυνητικά να μετριάσει σημαντικά το πρόβλημα της cold start, αξιοποιώντας τη γενική γνώση που έχουν αποκτήσει τα LLMs κατά την προεκπαίδευσή τους.

Παράλληλα, η αυξανόμενη κανονιστική πίεση (όπως ο Κανονισμός της Ευρωπαϊκής Ένωσης για την Τεχνητή Νοημοσύνη (EU AI Act)) αναμένεται να επηρεάσει καθοριστικά τον σχεδιασμό μελλοντικών συστημάτων συστάσεων, απαιτώντας ενισχυμένη διαφάνεια, ερμηνευσιμότητα και δυνατότητα ελέγχου των αλγοριθμικών αποφάσεων, ιδιαίτερα σε εφαρμογές υψηλού κινδύνου όπως αυτές που παρουσιάστηκαν στην Ενότητα 7.9 για τον τομέα της υγείας. Τέλος, η μετάβαση προς αρχιτεκτονικές on-device, που εκτελούν την εξατομίκευση τοπικά στη συσκευή του χρήστη χωρίς μεταφορά ευαίσθητων δεδομένων σε κεντρικούς εξυπηρετητές, αναμένεται να αποκτήσει αυξανόμενη σημασία ως απάντηση στις ανησυχίες

ιδιωτικότητας, εξισορροπώντας την ανάγκη για εξατομίκευση με το δικαίωμα στην ιδιωτικότητα.

10.4 Τελικές σκέψεις

Τα συστήματα συστάσεων έχουν ήδη μεταμορφώσει τον τρόπο με τον οποίο οι άνθρωποι ανακαλύπτουν προϊόντα, περιεχόμενο, εκπαίδευση και πολιτισμό. Η δυναμική ανάπτυξη των τεχνικών βαθιάς μάθησης, γραφικών νευρωνικών δικτύων και μεγάλων γλωσσικών μοντέλων υπόσχεται ότι τα επόμενα χρόνια θα αποφέρουν ακόμα πιο εξελιγμένα, ακριβή και εξατομικευμένα συστήματα. Ωστόσο, η τεχνολογική πρόοδος από μόνη της δεν αρκεί: η ανάπτυξη υπεύθυνων συστημάτων συστάσεων απαιτεί παράλληλα κριτική σκέψη για τις κοινωνικές τους επιπτώσεις, σεβασμό στην ιδιωτικότητα και τη δικαιοσύνη, και ανοιχτό διάλογο μεταξύ ερευνητών, κατασκευαστών, ρυθμιστικών αρχών και κοινωνίας.

Η παρούσα εργασία συνέβαλε στη χαρτογράφηση αυτού του πλούσιου και ταχύτατα εξελισσόμενου πεδίου. Η μελέτη εκατό επιστημονικών εργασιών από πολλαπλές κατευθύνσεις επέτρεψε μια ολοκληρωμένη εικόνα (τόσο της τεχνικής προόδου όσο και των ανοιχτών ζητημάτων) που ελπίζεται να αποτελέσει χρήσιμο σημείο αναφοράς για όσους ασχολούνται ή ενδιαφέρονται να ασχοληθούν με το πεδίο των ευφυών συστημάτων συστάσεων.

Παράρτημα Α: Αλγόριθμοι και Ψευδοκώδικας

Στο παράρτημα αυτό παρατίθενται οι αλγόριθμοι και ο ψευδοκώδικας των κυριότερων μεθόδων που αναλύθηκαν στην εργασία. Οι αλγόριθμοι παρουσιάζονται σε ψευδοκώδικα για μεγαλύτερη σαφήνεια και ανεξαρτησία από συγκεκριμένη γλώσσα προγραμματισμού.

A.1 Bayesian Personalized Ranking (BPR)

Ο αλγόριθμος BPR αποτελεί τη βάση εκπαίδευσης πολλών μοντέλων collaborative filtering, βελτιστοποιώντας απευθείας τη σχετική κατάταξη αντικειμένων αντί για την απόλυτη τιμή βαθμολογίας. Χρησιμοποιεί triplets (χρήστης, θετικό αντικείμενο, αρνητικό αντικείμενο) και στοχεύει στη μεγιστοποίηση της πιθανότητας ο χρήστης να προτιμά το θετικό έναντι του αρνητικού αντικειμένου.

```
Αλγόριθμος 1: Bayesian Personalized Ranking (BPR)
──────────────────────────────────────────────────

Είσοδος: Πίνακας αλληλεπιδράσεων R, αριθμός παραγόντων k,
         ρυθμός μάθησης (learning rate) α, regularization λ
Έξοδος:  Πίνακες P (χρήστες) και Q (αντικείμενα)

1: Αρχικοποίηση P, Q ~ N(0, σ²)
2: repeat
3:   για κάθε (u, i, j) ∈ DS  // i: θετικό, j: αρνητικό
4:     x̂_uij ← p_u · q_i − p_u · q_j
5:     δ ← 1 − σ(x̂_uij)        // sigmoid gradient
6:     p_u ← p_u + α · (δ · (q_i − q_j) − λ · p_u)
7:     q_i ← q_i + α · (δ · p_u − λ · q_i)
8:     q_j ← q_j + α · (−δ · p_u − λ · q_j)
9: until σύγκλιση
10: return P, Q
```

A.2 LightGCN - Γραφική Διάδοση και Πρόβλεψη

Το LightGCN απλοποιεί τη γραφική συνέλιξη αφαιρώντας μη γραμμικές ενεργοποιήσεις και μετασχηματισμούς χαρακτηριστικών. Η τελική αναπαράσταση κάθε κόμβου υπολογίζεται ως weighted sum των αναπαραστάσεων από όλα τα επίπεδα.

```
Αλγόριθμος 2: LightGCN - Propagation & Prediction
──────────────────────────────────────────────────────────
Είσοδος: Γράφος G=(U∪I, A), embeddings E⁰, επίπεδα L
Έξοδος:  Τελικά embeddings e_u, e_i

1: // Embedding Propagation
2: για l = 1 έως L:
3:   E^(l) ← D^(-½) · Ã · D^(-½) · E^(l-1)
   // Ã: κανονικοποιημένος πίνακας γειτνίασης
   // D: diagonal degree matrix

4: // Layer Combination (mean pooling)
5: E_final ← (1/(L+1)) · Σ_{l=0}^{L} E^(l)

6: // Prediction
7: ŷ_ui ← e_u^T · e_i

8: // BPR Loss
9: L_BPR = -Σ ln σ(ŷ_ui - ŷ_uj) + λ||E⁰||²
```

A.3 Self-supervised Graph Contrastive Learning (SGL)

Ο αλγόριθμος SGL δημιουργεί θετικά ζεύγη κόμβων μέσω τυχαίας επαύξησης γράφου και εκπαιδεύει μοντέλο χρησιμοποιώντας InfoNCE loss. Αντιμετωπίζει αποτελεσματικά την data sparsity και τη μεροληψία δημοτικότητας.

```
Αλγόριθμος 3: Self-supervised Graph Contrastive Learning (SGL)
──────────────────────────────────────────────────────────
Είσοδος: Γράφος G, augmentation operator Aug(·)
Έξοδος:  Ενισχυμένα embeddings

1: // Δημιουργία 2 augmented views
2: G' ← Aug(G)    // π.χ. random edge dropout p=0.1
3: G'' ← Aug(G)   // δεύτερο ανεξάρτητο view

4: // Εκπαίδευση GNN και στα 2 views
5: Z' ← GNN(G'),  Z'' ← GNN(G'')

6: // InfoNCE Contrastive Loss
7: L_CL = -Σ_u log [exp(sim(z'_u, z''_u)/τ) /
              Σ_{u'≠u} exp(sim(z'_u, z''_{u'})/τ)]
   // τ: temperature hyperparameter

8: // Συνολική Loss
9: L = L_BPR + λ · L_CL
10: Βελτιστοποίηση με Adam optimizer
```

A.4 FairDA - Fairness Dual-path Alignment

Ο αλγόριθμος FairDA αντιμετωπίζει ταυτόχρονα attribute bias και topology bias μέσω distillation από δίκαιο teacher μοντέλο και information bottleneck constraint.

```
Αλγόριθμος 4: FairDA - Dual-path Fairness Optimization
──────────────────────────────────────────────────────────────

Είσοδος: Γράφος G, ευαίσθητα χαρακτηριστικά S
Έξοδος:  Δίκαια embeddings χρηστών

// Βήμα 1: Fairness-Oriented Distillation (FOD)
1: G_attr ← G χωρίς attribute bias (αφαίρεση S)
2: G_topo ← G με ισορροπημένη συνδεσιμότητα
3: z_attr ← LightGCN(G_attr)  // Attribute expert
4: z_topo ← LightGCN(G_topo)  // Topology expert
5: z_teacher ← MLP([z_attr; z_topo])  // Fair teacher

// Βήμα 2: Student με Information Bottleneck
6: z_student ← LightGCN(G)  // Student σε πλήρη γράφο
7: L_IB ← MI(z_student; S) - β · MI(z_student; Y)
   // MI: Mutual Information, Y: recommendation labels

// Βήμα 3: DACR - Cross-group item alignment
8: Sim_matrix ← user co-occurrence similarity
9: για κάθε (i,j) ∈ similar_cross_group_pairs:
10:  w_ij ← dynamic_weight(i, j)
11:  L_DACR += w_ij · ||q_i - q_j||²

// Τελική βελτιστοποίηση
12: L_total = L_BPR + α·L_distill + β·L_IB + γ·L_DACR
```

A.5 Multi-Behavior Sequential Recommendation (MBSR)

Ο αλγόριθμος MBSR μοντελοποιεί ταυτόχρονα τη σειριακή φύση και την ετερογένεια των αλληλεπιδράσεων χρήστη, συνδυάζοντας Transformer encoders με cross-behavior attention.

```
Αλγόριθμος 5: Multi-Behavior Sequential Recommendation (MBSR)
──────────────────────────────────────────────────────────────

Είσοδος: Ακολουθίες {S^b_u} για κάθε τύπο συμπεριφοράς b
         Σύνολο συμπεριφορών B = {view, cart, fav, purchase}
Έξοδος:  Κατάταξη υποψηφίων αντικειμένων

1: // Κωδικοποίηση κάθε τύπου συμπεριφοράς
2: για κάθε b ∈ B:
3:   H^b ← Transformer_Encoder(S^b_u, pos_encoding)
4:   h^b_u ← MeanPool(H^b)  // βραχυπρόθεσμη πρόθεση

5: // Cross-behavior attention
6: για κάθε ζεύγος (b1, b2) ∈ B × B:
7:   α_{b1→b2} ← Attention(h^b1_u, h^b2_u)

8: // Μακροπρόθεσμες προτιμήσεις
9: h_long ← UserEmbedding(u)

10: // Συνένωση
11: h_final ← concat([h^purchase_u, Σ_b α_b·h^b_u, h_long])

12: // Πρόβλεψη επόμενου αντικειμένου
13: score(i) ← h_final · q_i
14: return top-K αντικείμενα βάσει score
```

A.6 Neural Collaborative Filtering (NeuMF)

Το NeuMF συνδυάζει μια GMF διαδρομή για γραμμική μοντελοποίηση και μια MLP διαδρομή για μη γραμμική σύλληψη, ξεπερνώντας σταθερά τη Matrix Factorization.

```
Αλγόριθμος 6: Neural Collaborative Filtering (NeuMF)
──────────────────────────────────────────────────────
Είσοδος: Χρήστης u, Αντικείμενο i
Έξοδος:  ŷ_ui ∈ [0,1] (πιθανότητα θετικής αλληλεπίδρασης)

// Generalized Matrix Factorization (GMF) path
1: p_u^GMF ← Embedding_GMF(u)   // k-dimensional
2: q_i^GMF ← Embedding_GMF(i)
3: φ^GMF ← p_u^GMF ⊙ q_i^GMF   // element-wise product

// MLP path
4: p_u^MLP ← Embedding_MLP(u)
5: q_i^MLP ← Embedding_MLP(i)
6: z^0 ← concat([p_u^MLP, q_i^MLP])
7: για l = 1 έως L:
8:   z^l ← ReLU(W^l · z^(l-1) + b^l)
9: φ^MLP ← z^L

// Fusion & Output
10: φ ← concat([φ^GMF, φ^MLP])
11: ŷ_ui ← σ(h^T · φ)  // sigmoid activation

// Loss (Binary Cross-Entropy)
12: L = -Σ [y_ui·log(ŷ_ui) + (1-y_ui)·log(1-ŷ_ui)]
```

A.7 Αλγόριθμος SASRec: Self-Attention Sequential RS

```
Αλγόριθμος 7: SASRec: Self-Attention Sequential Recommendation
──────────────────────────────────────────────────────────────
Είσοδος: Ακολουθία S_u = [i_1,...,i_t], παράμετροι L, h, d
Έξοδος:  score για κάθε υποψήφιο αντικείμενο

1: // Item Embedding + Positional Encoding
2: για k = 1 έως t:
3:   e_k ← Embedding(i_k) + PE(k)
4: E ← [e_1, ..., e_t]  // (t × d) matrix

5: // L επίπεδα Transformer
6: για l = 1 έως L:
7:   // Causal (unidirectional) Self-Attention
8:   Q = E·W_Q^l,  K = E·W_K^l,  V = E·W_V^l
9:   A = softmax(mask(QK^T / √d_k)) · V
10:  // Feed-Forward + Residual + LayerNorm
11:  F = ReLU(A·W_1^l + b_1^l)·W_2^l + b_2^l
12:  E ← LayerNorm(E + F)

13: // Prediction για επόμενο αντικείμενο
14: h_t ← E[t]  // τελευταία θέση
15: score(j) ← h_t · Embedding(j) για κάθε j
16: return top-K αντικείμενα βάσει score
```

A.8 Αλγόριθμος SimGCL: Simple Graph Contrastive Learning

```
Αλγόριθμος 8: SimGCL: Simple Graph Contrastive Learning
──────────────────────────────────────────────────────────

Είσοδος: Γράφος G, noise magnitude ε
Έξοδος:  Ενισχυμένα embeddings E

// SimGCL: αντί για structural augmentation, προσθέτει
// Gaussian noise απευθείας στα embeddings

1: // Forward pass LightGCN
2: E^(0) ← αρχικά embeddings
3: για l = 1 έως L:
4:   E^(l) ← D^(-½)ÃD^(-½)E^(l-1)
5: E ← (1/(L+1)) Σ E^(l)

6: // Δημιουργία 2 augmented views με noise
7: E' ← E + ε·Uniform(-1, 1) / ||E||  // view 1
8: E''← E + ε·Uniform(-1, 1) / ||E||  // view 2

9: // InfoNCE Loss
10: για κάθε χρήστη u:
11:   pos = exp(sim(e'_u, e''_u) / τ)
12:   neg = Σ_{u'≠u} exp(sim(e'_u, e''_u') / τ)
13:   L_CL += -log(pos / neg)

14: L = L_BPR + λ·L_CL
```

A.9 Αλγόριθμος GraphSAGE: Inductive Representation Learning

```
Αλγόριθμος 9: GraphSAGE: Inductive Representation Learning via Sampling and
Aggregation
──────────────────────────────────────────────────────────────────────────

Είσοδος: Γράφος G(V, E), χαρακτηριστικά κόμβων x_v για κάθε v∈V, αριθμός
         επιπέδων K, συναρτήσεις συνάθροισης AGGREGATE_k για k=1..K,
         μέγεθος δείγματος γειτονιάς S.

1:  h_v^0 ← x_v για κάθε v∈V
2:  για k = 1 έως K:
3:      για κάθε v∈V:
4:          N_S(v) ← τυχαίο δείγμα μεγέθους S από τους γείτονες του v
5:          h_{N(v)}^k ← AGGREGATE_k({h_u^{k-1}, ∀u∈N_S(v)})
6:          h_v^k ← σ(W^k · CONCAT(h_v^{k-1}, h_{N(v)}^k))
7:          h_v^k ← h_v^k / ‖h_v^k‖₂
8:  z_v ← h_v^K για κάθε v∈V

Έξοδος:  Τελικές αναπαραστάσεις κόμβων z_v
```

Σε αντίθεση με το LightGCN (Παράρτημα Α.2), το οποίο απαιτεί πρόσβαση σε ολόκληρο τον γράφο κατά την εκπαίδευση (transductive setting), το GraphSAGE μαθαίνει συναρτήσεις συνάθροισης γενικεύσιμες σε κόμβους που δεν είχαν παρατηρηθεί κατά την εκπαίδευση (inductive setting), καθιστώντας το ιδιαίτερα κατάλληλο για πλατφόρμες όπου προστίθενται συνεχώς νέοι χρήστες ή αντικείμενα, συνδέοντάς το άμεσα με τις τεχνικές δειγματοληψίας γειτονιάς που αναλύθηκαν στην Ενότητα 5.14.

Παράρτημα Β: Αναλυτικοί Πίνακες Αποτελεσμάτων

Στο παράρτημα αυτό παρατίθενται αναλυτικοί συγκριτικοί πίνακες αποτελεσμάτων, δεδομένων και μοντέλων που αναφέρθηκαν στα κεφάλαια της εργασίας.

Β.1 Σύγκριση Μοντέλων Sequential Recommendation

Ο Πίνακας Β.1 παρουσιάζει τα κυριότερα μοντέλα σειριακής σύστασης με τα αντίστοιχα αποτελέσματα σε benchmark datasets. Τα αποτελέσματα αντλήθηκαν από τα papers που αναλύθηκαν.

Πίνακας Β.1: Σύγκριση μοντέλων σειριακής σύστασης

Μοντέλο	Έτος	Βάση	Dataset	NDCG@10	Βασική Καινοτομία
GRU4Rec	2016	GRU	RSC15	—	Πρώτο RNN για session RS
HRNN	2017	RNN	MovieLens	—	Ιεραρχική μοντελοποίηση
Caser	2018	CNN	Gowalla	0.3162	Οριζόντια+κάθετα φίλτρα
SASRec	2018	Transformer	ML-1M	0.5857	Unidirectional self-attention
BERT4Rec	2019	BERT	ML-1M	0.5943	Bidirectional + masked prediction
SR-GNN	2019	GNN+GRU	Diginetica	0.3560	Session graph + Gated GNN
LightGCN	2020	GCN	Gowalla	0.1254	Μη-σειριακό CF/GNN baseline (γενικού σκοπού, όχι sequential)
CL4Rec	2022	SSL+CF	ML-1M	0.6023	Contrastive Learning για RS
SGL	2021	GCL	Yelp2018	0.0593	Self-supervised Graph Learning
SimGCL	2022	GCL	Yelp2018	0.0695	Noise-based augmentation
BSARec	2024	Transformer	ML-1M	0.6218	Bandwidth self-attention

Β.2 Ποσοτικά Αποτελέσματα GNN Μοντέλων

Ο Πίνακας Β.2 παρουσιάζει τα ποσοτικά αποτελέσματα των κυριότερων GNN μοντέλων σε πραγματικά benchmark datasets, αναδεικνύοντας την εξελικτική πορεία από το NGCF στο LightGCN και τα GCL μοντέλα.

Πίνακας Β.2: Ποσοτικά αποτελέσματα GNN μοντέλων

Μοντέλο	Dataset	Recall@20	NDCG@20	Precision @20	Τύπος Γράφου
NGCF	Gowalla	0.1547	0.0639	—	Bipartite user-item
LightGCN	Gowalla	0.1830	0.0761	—	Bipartite user-item
SGL	Gowalla	0.1954	0.0795	—	Bipartite + augmented
SimGCL	Gowalla	0.2028	0.0861	—	Bipartite + noise
HDHP (HGCN)	MovieLens 100K	0.8876	0.4201	0.5918	Heterogeneous Hypergraph
KGCN	MovieLens-1M	—	—	0.976 (AUC)	Knowledge Graph
MycGNN	RetailRocket	0.2055	0.0827	—	Dynamic Bipartite

Β.3 Αποτελέσματα ανά Τομέα Εφαρμογής

Ο Πίνακας Β.3 συγκεντρώνει τα σημαντικότερα ποσοτικά αποτελέσματα από διαφορετικούς τομείς εφαρμογής, επιτρέποντας οριζόντια σύγκριση της αποτελεσματικότητας των μεθόδων σε ετερογενή περιβάλλοντα.

Πίνακας Β.3: Αποτελέσματα μοντέλων ανά τομέα εφαρμογής

Τομέας	Μοντέλο	Dataset	Μετρική	Αποτέλεσμα
E-commerce	CNN + LSTM + Attention	Alibaba	Accuracy	89% (vs CF: 72%, MF: 78%)
E-commerce	GNN (Heterogeneous)	Taobao	Recall@10	Σημαντική υπεροχή vs CF
Ξενοδοχεία	DeepMF + Sentiment	Hotel Reviews	RMSE	0.10 (Sentiment Acc: 88.63%)
Ταξίδια	ResNetMF (MCRS)	Τουριστικά δεδομένα	RMSE/MAE	Ανώτερο vs CF, MF, DNN baselines
Ταξίδια	DRL + GNN	Τουριστικά δεδομένα	Accuracy/Satisfaction	0.85 / 4.6 (από 5)
Εκπαίδευση	CBO-FFNN	Library (42K χρήστες)	Accuracy/F-value	91.4% / 95.7%
Εκπαίδευση	DNN + RL	Πανεπιστημιακό σύστημα	Μαθησιακά αποτελέσματα	+20% μετά real-time προσαρμογή
Τρόφιμα	FRMADHG (Hypergraph)	Food.com (32K χρήστες)	Precision@10	+19.8% vs CF baselines ($p<0.001$)
Τρόφιμα	FRMADHG (Hypergraph)	Allrecipes (25K χρήστες)	Recall@10	+18.7% vs CF baselines ($p<0.001$)
Κινηματογράφος	RBM + kNN CF	MovieLens-1M	RMSE/MAE	0.8736 / 0.6854
Κοιν. Δίκτυα	DRLGR (Graph Rewiring)	Video/News dataset	Radicalization Score	Στατιστικά σημαντική μείωση

Β.4 Σύγκριση Μετρικών Αξιολόγησης ανά Κατηγορία Μεθόδου

Ο Πίνακας Β.4 παρέχει ενδεικτικές τιμές των κυριότερων μετρικών αξιολόγησης ανά κατηγορία μεθόδου, βασισμένες σε αντιπροσωπευτικά πειράματα της βιβλιογραφίας. Οι τιμές αποτελούν εκτιμήσεις κεντρικής τάσης βάσει πολλαπλών πειραμάτων.

Πίνακας Β.4: Ενδεικτικές τιμές μετρικών ανά κατηγορία μεθόδου

Κατηγορία Μεθόδου	RMSE (↓)	MAE (↓)	P@10 (↑)	R@10 (↑)	Παρατήρηση
Memory-based CF	~1.02	~0.81	~0.62	~0.53	Baseline απόδοση
Matrix Factorization (SVD)	~0.91	~0.72	~0.70	~0.61	Σημαντική βελτίωση vs CF
NCF / NeuMF	~0.87	~0.69	~0.76	~0.67	Μη γραμμική σύλληψη
RNN/LSTM (GRU4Rec)	~0.86	~0.68	~0.78	~0.70	Sequential modeling
Transformer (SASRec)	~0.84	~0.66	~0.81	~0.73	Μακρές εξαρτήσεις
GCN / LightGCN	~0.83	~0.65	~0.83	~0.76	Γραφική πληροφορία
GCL / SimGCL	~0.82	~0.64	~0.86	~0.79	SOTA + fairness + diversity
Hybrid DL + MF	~0.10*	—	~0.89	—	*DeepMF+Sentiment (hotel)

Β.5 Αναλυτικά Στατιστικά Benchmark Datasets

Ο Πίνακας Β.5 παρέχει αναλυτικά στατιστικά για τα κυριότερα benchmark datasets που χρησιμοποιούνται στην αξιολόγηση των συστημάτων συστάσεων, συμπεριλαμβανομένης της πυκνότητας δεδομένων: ένα κρίσιμο χαρακτηριστικό που επηρεάζει σημαντικά την απόδοση των αλγορίθμων.

Πίνακας Β.5: Αναλυτικά στατιστικά benchmark datasets

Dataset	Χρήστες	Αντικείμενα	Αλληλεπιδράσεις	Πυκνότητα	Μετρική Αναφοράς
MovieLens 100K	943	1.682	100.000	6.30%	RMSE, MAE
MovieLens 1M	6.040	3.952	1.000.209	4.19%	RMSE, MAE, NDCG
MovieLens 20M	138.493	27.278	20.000.263	0.53%	Precision, Recall
Amazon Beauty	22.363	12.101	198.502	0.07%	NDCG@10, HR@10
Yelp2018	31.668	38.048	1.561.406	0.13%	Recall@20, NDCG@20
Gowalla	29.858	40.981	1.027.370	0.08%	Recall@20, NDCG@20
Taobao (MBSR)	48.749	39.493	2.641.225	0.14%	Recall@10, NDCG@10
Beibei (MBSR)	21.716	7.977	3.338.118	1.93%	Recall@10, NDCG@10
RetailRocket	1.407.580	235.061	2.764.543	0.00083%	Recall@20, MRR@20
Food.com	32.000	18.500	1.100.000+	0.19%	Precision@K, Recall@K

Β.6 Σύγκριση Μοντέλων Multi-Behavior Sequential Recommendation

Ο Πίνακας Β.6 παρουσιάζει τα κυριότερα μοντέλα MBSR με τα αποτελέσματά τους σε πραγματικά e-commerce datasets, αναδεικνύοντας την εξελικτική πορεία από απλά CF-based μοντέλα έως σύνθετες GNN και Transformer αρχιτεκτονικές.

Πίνακας Β.6: Σύγκριση μοντέλων Multi-Behavior Sequential Recommendation

Μοντέλο	Αρχιτεκτονική	Dataset	Recall@10	NDCG@10	Χειρισμός Πολλαπλών Συμπεριφορών
EHCF	CF-based	Taobao	0.0735	0.0459	Transfer Learning μεταξύ συμπεριφορών
MBGCN	GNN	Taobao	0.0781	0.0492	Ξεχωριστός γράφος ανά συμπεριφορά
CML	GNN+Attention	Taobao	0.0830	0.0521	Cross-behavior contrastive learning
GHCF	Hypergraph	Beibei	0.1843	0.1139	Hypergraph για ομαδικές σχέσεις
HMGCR	GNN+Transformer	Taobao	0.0892	0.0561	Ιεραρχική μοντελοποίηση

Β.7 Τεχνικές Αντιμετώπισης Προκλήσεων

Ο Πίνακας Β.7 παρέχει λεπτομερή αντιστοίχιση μεταξύ των κυριότερων προκλήσεων, της επίδρασής τους στην απόδοση, του επιπέδου σοβαρότητάς τους και των τεχνικών αντιμετώπισής τους, βάσει της βιβλιογραφικής ανάλυσης.

Πίνακας Β.7: Τεχνικές αντιμετώπισης προκλήσεων

Πρόκληση	Επίδραση στην Απόδοση	Επίπεδο Σοβαρότητας	Τεχνικές Αντιμετώπισης
Data Sparsity	RMSE ↑ 15-30%	Πολύ Υψηλό	GCL, SSL, Data Augmentation, Knowledge Graph
Cold Start (Χρήστη)	Recall ↓ έως 60%	Πολύ Υψηλό	Meta-Learning, Conversational RS, LLM, Content-based init
Cold Start (Αντικειμένου)	Recall ↓ έως 50%	Υψηλό	Knowledge Graph, Transfer Learning, Content features
Popularity Bias	Diversity ↓ 30-45%	Υψηλό	IPS, FairDA, MycGNN, Inverse Popularity Weighting
Over-smoothing (GNN)	Accuracy ↓ με depth>4	Μέτριο	LightGCN, Residual connections, Graph Transformers
Shilling Attacks	Accuracy ↓ 10-25%	Υψηλό	Robust Training, Anomaly Detection, Certified Defense
Ιδιωτικότητα	Νομική μη συμμόρφωση	Κρίσιμο	Federated Learning, Differential Privacy, On-device RS
Κλιμακωσιμότητα	Latency ↑ εκθετικά	Υψηλό	Two-stage RS, ANN Search, Distributed Training, LightGCN
Filter Bubble	Radicalization ↑	Κρίσιμο	DRLGR, Diversity-aware, Re-ranking, Fairness constraints

Β.8 Ενδεικτική Σύγκριση Υπολογιστικού Κόστους Εκπαίδευσης

Ο Πίνακας Β.8 παρουσιάζει ενδεικτική σύγκριση του σχετικού υπολογιστικού κόστους εκπαίδευσης για τις κυριότερες οικογένειες μοντέλων που αναλύθηκαν στην εργασία, βάσει της θεωρητικής πολυπλοκότητας του Παραρτήματος Γ (Ενότητα Γ.3) και ποιοτικών αναφορών της βιβλιογραφίας. Οι τιμές εκφράζονται ως σχετικός δείκτης (1× = μέθοδος αναφοράς) και όχι ως απόλυτος χρόνος, καθώς ο τελευταίος εξαρτάται σημαντικά από το υλικό εκτέλεσης (CPU/GPU) και το μέγεθος του dataset.

Πίνακας Β.8: Ενδεικτική σύγκριση σχετικού υπολογιστικού κόστους ανά οικογένεια μεθόδων

Μέθοδος	Σχετικό Κόστος Εκπαίδευσης	Σχετικό Κόστος Inference	Απαιτήσεις Μνήμης
User/Item-based CF	1× (αναφορά)	Χαμηλό	Χαμηλές
Matrix Factorization (ALS/SVD)	2-3×	Πολύ χαμηλό	Χαμηλές
NCF / NeuMF	8-12×	Μέτριο	Μέτριες
LightGCN (3 επίπεδα)	15-20×	Μέτριο	Μέτριο / Υψηλό
SASRec (Transformer)	25-35×	Υψηλό	Υψηλές
GCL / SGL	30-40×	Μέτριο-Υψηλό	Υψηλές

Β.9 Τύποι Επιθέσεων και Μηχανισμοί Άμυνας

Ο Πίνακας Β.9 συνοψίζει τους κύριους τύπους ανταγωνιστικών επιθέσεων που αναλύθηκαν στην Ενότητα 8.7.1, μαζί με τους αντίστοιχους στόχους τους και τις κυριότερες τεχνικές άμυνας που αναφέρονται στη βιβλιογραφία.

Πίνακας Β.9: Τύποι επιθέσεων σε συστήματα συστάσεων και μηχανισμοί άμυνας

Τύπος Επίθεσης	Στόχος	Κύρια Άμυνα
Shilling (μεμονωμένος)	Ενίσχυση/υποβάθμιση αντικειμένου	Ανίχνευση ανωμαλιών προφίλ
Shilling groups	Συντονισμένη χειραγώγηση κατάταξης	Ανάλυση συσχέτισης λογαριασμών
Adversarial noise (embeddings)	Παραπλάνηση μοντέλου πρόβλεψης	Adversarial training
Data poisoning	Υποβάθμιση συνολικής απόδοσης	Robust loss functions

Β.10 Σύγκριση Αναδυόμενων Τομέων Εφαρμογής

Ο Πίνακας Β.10 συγκρίνει τους πρόσθετους τομείς εφαρμογής που αναλύθηκαν στις Ενότητες 7.1, 7.3, 7.9 και 7.10, αναδεικνύοντας το ιδιαίτερο χαρακτηριστικό κάθε τομέα και την κυρίαρχη τεχνική προσέγγιση που υιοθετείται στη σχετική βιβλιογραφία.

Πίνακας Β.10: Σύγκριση αναδυόμενων τομέων εφαρμογής συστημάτων συστάσεων

Τομέας	Ιδιαίτερο Χαρακτηριστικό	Κύρια Τεχνική
Υγεία/Φαρμακευτική	Κρισιμότητα αποφάσεων, ανάγκη ερμηνευσιμότητας	Ετερογενή GNN + κλινική επικύρωση
Γεωργία	Αντικειμενικά κριτήρια απόδοσης	IoT δεδομένα + context-aware RS
Εκπαίδευση	Προσαρμογή στον ρυθμό μάθησης	Γνωστική διάγνωση + adaptive learning

Τομέας	Ιδιαίτερο Χαρακτηριστικό	Κύρια Τεχνική
Κοινωνικό e-commerce	Επίδραση κοινωνικού δικτύου	Ετερογενές GNN κοινωνικού γράφου

Παράρτημα Γ: Μαθηματικές αποδείξεις και αναλύσεις

Στο παράρτημα αυτό παρατίθενται αναλυτικές μαθηματικές αποδείξεις και αναλύσεις που υποστηρίζουν το θεωρητικό υπόβαθρο της εργασίας.

Γ.1 Απόδειξη σύγκλισης BPR με SGD

Η συνάρτηση κόστους BPR ορίζεται ως: $L_{BPR} = -\Sigma_{(u,i,j)\in DS} \ln \sigma(\hat{r}_{ui} - \hat{r}_{uj}) + (\lambda/2)||\Theta||^2$, όπου σ() η συνάρτηση sigmoid. Για να αποδειχθεί η σύγκλιση, αρκεί να δειχθεί ότι η συνάρτηση είναι φραγμένη από κάτω και ότι οι κλίσεις είναι φραγμένες.

Η κλίση ως προς p_u είναι: $\partial L/\partial p_u = -\Sigma_{(u,i,j)} (1-\sigma(x_{uij}))(q_i - q_j) + \lambda p_u$, όπου $x_{uij} = \hat{r}_{ui} - \hat{r}_{uj}$. Επειδή $\sigma(x) \in (0,1)$ για κάθε x, ο όρος $(1-\sigma(x_{uij})) \in (0,1)$, άρα η κλίση είναι φραγμένη. Με βήμα μάθησης α που ικανοποιεί τις συνθήκες Armijo, η μέθοδος SGD συγκλίνει σε τοπικό ελάχιστο.

Σημειώνεται ότι η συνάρτηση κόστους BPR δεν είναι κυρτή ως προς το σύνολο των παραμέτρων $\Theta = \{P, Q\}$, λόγω του διγραμμικού όρου $\hat{r}_{ui} = p_u \cdot q_i^t$. Συνεπώς, η απόδειξη σύγκλισης που παρουσιάστηκε εγγυάται σύγκλιση σε τοπικό, όχι απαραίτητα καθολικό, ελάχιστο. Στην πράξη, ωστόσο, η εμπειρική απόδοση των μοντέλων BPR παραμένει ισχυρή, καθώς η επιφάνεια απώλειας (loss landscape) διγραμμικών μοντέλων χαμηλής τάξης (low-rank) έχει αποδειχθεί ότι διαθέτει σχετικά «καλοήθη» γεωμετρία, με τα τοπικά ελάχιστα να βρίσκονται συνήθως κοντά στο καθολικό ελάχιστο.

Γ.2 Ιδιότητες κανονικοποίησης (regularization) LightGCN

Στο LightGCN, ο κανονικοποιημένος πίνακας γειτνίασης ορίζεται ως $\tilde{A} = D^{(-½)} \cdot A \cdot D^{(-½)}$, όπου D είναι ο διαγώνιος πίνακας βαθμού (degree matrix). Οι ιδιοτιμές του $\tilde{A}$ βρίσκονται στο διάστημα [-1, 1], γεγονός που εξασφαλίζει τη σταθερότητα της διάδοσης. Συγκεκριμένα, ο κανονικοποιημένος πίνακας $\tilde{A}$ είναι όμοιος με έναν συμμετρικό πίνακα, του οποίου το φάσμα (οι ιδιοτιμές) ανήκει στο διάστημα [-1, 1]· οι βαθμοί d_u, d_i λειτουργούν ως όροι κανονικοποίησης ανά κόμβο και όχι ως ενιαίοι διαιρέτες των ιδιοτιμών του A. Αυτό σημαίνει ότι τα σήματα δεν ενισχύονται κατά τη διάδοση, αποφεύγοντας το πρόβλημα της εκρηκτικής κλίσης.

Η ιδιότητα αυτή έχει, ωστόσο, και μια σημαντική παρενέργεια: καθώς ο αριθμός των επιπέδων διάδοσης L αυξάνεται, η επανειλημμένη εφαρμογή του $\tilde{A}$ οδηγεί τις αναπαραστάσεις των κόμβων να συγκλίνουν προς τον κυρίαρχο ιδιοχώρο του πίνακα, ο οποίος αντιστοιχεί στη σταθερή κατανομή (stationary distribution) του γράφου. Το φαινόμενο αυτό, γνωστό ως over-smoothing, αναλύεται λεπτομερέστερα στην Ενότητα 5.12, και αποτελεί τη θεωρητική αιτία για την οποία τα βαθύτερα GNN μοντέλα δεν αποδίδουν πάντα καλύτερα από τα πιο «ρηχά» (shallow) μοντέλα στο πλαίσιο των συστημάτων συστάσεων.

Γ.3 Ανάλυση πολυπλοκότητας κυριότερων αλγορίθμων

Πίνακας Γ.1: Ανάλυση πολυπλοκότητας χώρου, εκπαίδευσης και inference ανά αλγόριθμο

Αλγόριθμος	Χώρος	Εκπαίδευση	Inference
User-based CF	$O(\|U\|\times\|I\|)$	$O(\|U\|^2\times\|I\|)$	$O(\|U\|\times k)$
Item-based CF	$O(\|I\|^2)$	$O(\|I\|^2\times\|U\|)$	$O(\|I\|\times k)$
Matrix Factorization	$O((\|U\|+\|I\|)\times k)$	$O(\|R\|\times k)$	$O(k)$
NCF / NeuMF	$O((\|U\|+\|I\|)\times k)$	$O(\|R\|\times k^2)$	$O(k^2)$
LightGCN (L επίπεδα)	$O((\|U\|+\|I\|)\times k)$	$O(L\times\|R\|\times k)$	$O(L\times k)$
SASRec (μήκος n)	$O((\|U\|+\|I\|)\times k)$	$O(n^2\times d\times\|U\|)$	$O(n^2\times d)$
GCL / SGL	$O((\|U\|+\|I\|)\times k)$	$O(2\times L\times\|R\|\times k)$	$O(L\times k)$

Γ.4 Παράγωγος της συνάρτησης απώλειας στο Neural Collaborative Filtering

Στο μοντέλο NeuMF (Ενότητα 4.1, Σχήμα 4.1), η εκπαίδευση πραγματοποιείται συνήθως με ελαχιστοποίηση της διασταυρωτικής εντροπίας (binary cross-entropy) μεταξύ της προβλεπόμενης πιθανότητας αλληλεπίδρασης $\hat{y}_{ui}$ και της πραγματικής ετικέτας $y_{ui} \in \{0,1\}$: $L = -\Sigma_{(u,i)\in O \cup O^-} [y_{ui} \cdot \ln \hat{y}_{ui} + (1-y_{ui}) \cdot \ln(1-\hat{y}_{ui})]$, όπου O το σύνολο των παρατηρούμενων θετικών αλληλεπιδράσεων και O^- το σύνολο των αρνητικών δειγμάτων που προκύπτουν από αρνητική δειγματοληψία (Ενότητα 2.9).

Δεδομένου ότι $\hat{y}_{ui} = \sigma(h^t \cdot \varphi)$, όπου $\sigma()$ η σιγμοειδής συνάρτηση και φ το διάνυσμα fusion των κλάδων GMF και MLP, η κλίση της απώλειας ως προς την έξοδο πριν τη σιγμοειδή (logit) λαμβάνει την απλή μορφή: $\partial L / \partial (h^t \cdot \varphi) = \hat{y}_{ui} - y_{ui}$. Αυτή η ιδιότητα (γνωστή ιδιότητα της σιγμοειδούς συνάρτησης σε συνδυασμό με τη διασταυρωτική εντροπία) εξηγεί εν μέρει τη σταθερότητα της εκπαίδευσης μοντέλων τύπου NeuMF, καθώς η κλίση είναι ανάλογη του σφάλματος πρόβλεψης και παραμένει πάντα φραγμένη στο διάστημα [-1, 1], αποφεύγοντας το πρόβλημα της εξαφανιζόμενης κλίσης (vanishing gradient) που μπορεί να εμφανιστεί σε βαθύτερα δίκτυα με μη κατάλληλη αρχικοποίηση.

Γ.5 Παραγωγή του κάτω φράγματος ELBO για VAE

Στο μοντέλο Variational Autoencoder που παρουσιάστηκε στην Ενότητα 4.13, ο στόχος είναι η μεγιστοποίηση της λογαριθμικής πιθανοφάνειας $\ln p_\theta(x)$ των παρατηρούμενων δεδομένων x (π.χ. του διανύσματος αλληλεπιδράσεων ενός χρήστη). Δεδομένου ότι ο άμεσος υπολογισμός του $p_\theta(x) = \int p_\theta(x|z) \cdot p(z)\, dz$ είναι αδιαπέραστος (intractable) λόγω της ολοκλήρωσης πάνω σε όλο τον latent space z, εισάγεται μια προσεγγιστική κατανομή $q_\varphi(z|x)$ και αξιοποιείται η ανισότητα Jensen για την εξαγωγή ενός υπολογιστικά εφικτού κάτω φράγματος.

Αναλυτικά: $\ln p_\theta(x) = \ln \int p_\theta(x,z)\, dz = \ln \int q_\varphi(z|x) \cdot [p_\theta(x,z)/q_\varphi(z|x)]\, dz = \ln E_{q_\varphi(z|x)}[p_\theta(x,z)/q_\varphi(z|x)]$. Εφαρμόζοντας την ανισότητα Jensen ($\ln E[X] \geq E[\ln X]$ για κοίλη συνάρτηση όπως ο λογάριθμος), προκύπτει: $\ln p_\theta(x) \geq E_{q_\varphi(z|x)}[\ln(p_\theta(x,z)/q_\varphi(z|x))] = E_{q_\varphi(z|x)}[\ln p_\theta(x|z)] - D_{KL}(q_\varphi(z|x) \,\|\, p(z)) = \text{ELBO}$. Το κενό (gap) μεταξύ της πραγματικής λογαριθμικής πιθανοφάνειας και του ELBO ισούται ακριβώς με $D_{KL}(q_\varphi(z|x) \,\|\, p_\theta(z|x))$, δηλαδή την απόκλιση της προσεγγιστικής κατανομής από την πραγματική εκ των υστέρων (posterior) κατανομή: συνεπώς η μεγιστοποίηση του ELBO ισοδυναμεί με ταυτόχρονη μεγιστοποίηση της πιθανοφάνειας και ελαχιστοποίηση αυτής της απόκλισης.

Η πρακτική σημασία αυτής της παραγώγου είναι ότι το ELBO, σε αντίθεση με την αρχική πιθανοφάνεια, μπορεί να υπολογιστεί και να βελτιστοποιηθεί μέσω τυπικών μεθόδων gradient descent, αρκεί η κατανομή $q_\varphi(z|x)$ να επιτρέπει δειγματοληψία μέσω του τεχνάσματος επαναπαραμετροποίησης που αναφέρθηκε στην Ενότητα 4.13, καθιστώντας έτσι εφικτή την εκπαίδευση VAE-based μοντέλων σύστασης με τυπικές αρχιτεκτονικές νευρωνικών δικτύων και αλγόριθμους οπισθοδιάδοσης.

Πίνακας Ορολογίας

Στον ακόλουθο πίνακα φαίνεται η αντιστοίχιση μεταξύ των ξενόγλωσσων επιστημονικών όρων που χρησιμοποιούνται στην παρούσα εργασία και των ελληνικών αποδόσεών τους.

Πίνακας Ορολογίας: Αντιστοίχιση ξενόγλωσσων και ελληνικών επιστημονικών όρων

Ξενόγλωσσος Όρος	Ελληνικός Όρος
Recommendation System (RS)	Σύστημα Συστάσεων
Recommender System	Σύστημα Σύστασης
Collaborative Filtering (CF)	Συνεργατική Διήθηση
Content-based Filtering	Διήθηση Βάσει Περιεχομένου
Hybrid Recommendation	Υβριδική Σύσταση
Matrix Factorization (MF)	Παραγοντοποίηση Πίνακα
Singular Value Decomposition (SVD)	Αποσύνθεση Ιδιαζουσών Τιμών
Latent Factor	Λανθάνων Παράγοντας
Embedding	Διανυσματική Αναπαράσταση
User-Item Interaction Matrix	Πίνακας Αλληλεπίδρασης Χρήστη-Αντικειμένου
Utility Matrix	Πίνακας Χρησιμότητας
Explicit Feedback	Ρητή Ανάδραση
Implicit Feedback	Έμμεση Ανάδραση
Data Sparsity	Αραιότητα Δεδομένων
Cold Start	Ψυχρή Εκκίνηση
Popularity Bias	Μεροληψία Δημοτικότητας
Scalability	Κλιμακωσιμότητα
Deep Learning	Βαθιά Μάθηση
Neural Network	Νευρωνικό Δίκτυο
Multi-Layer Perceptron (MLP)	Πολυεπίπεδο Perceptron
Neural Collaborative Filtering (NCF)	Νευρωνική Συνεργατική Διήθηση
Recurrent Neural Network (RNN)	Αναδρομικό Νευρωνικό Δίκτυο
Long Short-Term Memory (LSTM)	Μακροπρόθεσμη Βραχυπρόθεσμη Μνήμη
Gated Recurrent Unit (GRU)	Πύλη Αναδρομικής Μονάδας

Convolutional Neural Network (CNN)	Νευρωνικό Δίκτυο Συνέλιξης
Autoencoder (AE)	Αυτόματο Κωδικοποιητικό
Transformer	Μετασχηματιστής
Self-Attention Mechanism	Μηχανισμός Αυτο-Προσοχής
Generative Adversarial Network (GAN)	Γεννητικό Ανταγωνιστικό Δίκτυο
Self-Supervised Learning (SSL)	Αυτο-Επιβλεπόμενη Μάθηση
Contrastive Learning	Αντιπαραθετική Μάθηση
Graph Neural Network (GNN)	Γραφικό Νευρωνικό Δίκτυο
Graph Convolutional Network (GCN)	Γραφικό Συνελικτικό Δίκτυο
Graph Attention Network (GAT)	Γραφικό Δίκτυο Προσοχής
Message Passing	Μεταβίβαση Μηνυμάτων
Over-smoothing	Υπερ-εξομάλυνση
Bipartite Graph	Δίγραφος
Knowledge Graph (KG)	Γνωσιακός Γράφος
Hypergraph	Υπεργράφος
Sequential Recommendation	Σειριακή Σύσταση
Session-based Recommendation	Σύσταση Βάσει Συνόδου
Context-aware Recommendation	Σύσταση Ευαίσθητη στο Πλαίσιο
Multi-behavior Recommendation	Σύσταση Πολλαπλής Συμπεριφοράς
Trust-based Recommendation	Σύσταση Βάσει Εμπιστοσύνης
Explainable Recommendation	Ερμηνεύσιμη Σύσταση
Fairness-aware Recommendation	Σύσταση με Επίγνωση Δικαιοσύνης
Conversational Recommender System	Σύστημα Σύστασης Μέσω Διαλόγου
Large Language Model (LLM)	Μεγάλο Γλωσσικό Μοντέλο
Federated Learning	Ομοσπονδιακή Μάθηση
Differential Privacy	Διαφορική Ιδιωτικότητα
Filter Bubble	Φυσαλίδα Φίλτρων
Echo Chamber	Θάλαμος Ηχώς
Information Overload	Πληροφοριακή Υπερφόρτωση
Precision@K	Ακρίβεια@K
Recall@K	Ανάκληση@K
Normalized Discounted Cumulative Gain (NDCG)	Κανονικοποιημένο Κέρδος Αθροιστικής Κατάταξης
Root Mean Squared Error (RMSE)	Τετραγωνική Ρίζα Μέσου Τετραγωνικού Σφάλματος
Mean Absolute Error (MAE)	Μέση Απόλυτη Απόκλιση
Reinforcement Learning (RL)	Ενισχυτική Μάθηση
Transfer Learning	Μεταφορά Μάθησης
Data Augmentation	Επαύξηση Δεδομένων

Πίνακας Συντμήσεων – Αρκτικόλεξων

Ο ακόλουθος πίνακας περιλαμβάνει τις συντομογραφίες και τα αρκτικόλεξα που χρησιμοποιούνται στην παρούσα εργασία.

Συντομογραφία	Πλήρης Ονομασία
AE	Autoencoder
ALS	Alternating Least Squares
ANN	Approximate Nearest Neighbor
BERT	Bidirectional Encoder Representations from Transformers
BPR	Bayesian Personalized Ranking
CARS	Context-Aware Recommender Systems
CF	Collaborative Filtering
CNN	Convolutional Neural Network
CRS	Conversational Recommender System
DP	Differential Privacy
DL	Deep Learning
DNN	Deep Neural Network
ECTS	European Credit Transfer and Accumulation System
FL	Federated Learning
GAT	Graph Attention Network
GCL	Graph Contrastive Learning
GCN	Graph Convolutional Network
GDPR	General Data Protection Regulation
GNN	Graph Neural Network
GRU	Gated Recurrent Unit
KG	Knowledge Graph
LLM	Large Language Model
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MBSR	Multi-Behavior Sequential Recommendation
MF	Matrix Factorization
MLP	Multi-Layer Perceptron

NCF	Neural Collaborative Filtering
NDCG	Normalized Discounted Cumulative Gain
NGCF	Neural Graph Collaborative Filtering
PMF	Probabilistic Matrix Factorization
RBM	Restricted Boltzmann Machine
RL	Reinforcement Learning
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RS	Recommendation System / Recommender System
SGD	Stochastic Gradient Descent
SSL	Self-Supervised Learning
SVD	Singular Value Decomposition
TF-IDF	Term Frequency – Inverse Document Frequency

Βιβλιογραφία

Οι παρακάτω βιβλιογραφικές αναφορές παρατίθενται με αλφαβητική σειρά βάσει του επωνύμου του πρώτου συγγραφέα, σύμφωνα με το πρότυπο του Τμήματος Πληροφορικής του Πανεπιστημίου Πειραιώς.

Aarif, M. et al. (2026). Deep neural collaborative filtering model for personalized travel recommendation. Scientific Reports (Nature Research). https://doi.org/10.1038/s41598-026

Abakarim, Y. et al. (2026). Emerging trends of recommender system for e-commerce: a comprehensive review. Springer Science and Business Media BV.

Akdim, K. et al. (2026). Trust in recommender systems: A survey. Expert Systems with Applications.

Alam, M. & Ahmed, F. (2025). Dual-module wider and deeper stochastic gradient descent and dropout based dense neural network for movie recommendation. Scientific Reports (Nature Research).

Alkhrijah, Y. et al. (2026). Adaptive dynamic hypergraph learning for ingredient-aware food recommendation. Scientific Reports (Nature Research).

Bahi, M. et al. (2026). MycGNN: Enhancing recommendation diversity in e-commerce through mycelium-inspired graph neural network. Springer.

Berjawi, R. et al. (2025). Mitigating radicalization in recommender systems by rewiring. Online Social Networks and Media.

Bian, X. & Chang, Y. (2026). Design and optimization of education informatization model and intelligent recommendation system based on student perception. Springer Nature.

Chafiki, M. E. A. et al. (2026). A systematic review of advancements in context-aware recommendation systems. Knowledge and Information Systems, 68(11). https://doi.org/10.1007/s10115-025-02627-8

Chen, J. et al. (2026). A survey on multi-behavior sequential recommendation. Science China Information Sciences, 69(3), 131101. https://doi.org/10.1007/s11432-024-4568-7

Chen, X. et al. (2022). Deep-learning-based hashtag recommendation system for multimodal content. Information Sciences.

Da'u, A. et al. (2021). An adaptive deep learning method for item recommendation system. Knowledge-Based Systems, 213, 106681.

Darraz, N. et al. (2025). Advancing recommendation systems with DeepMF and hybrid sentiment

analysis: Deep learning and Lexicon-based integration. Expert Systems With Applications, 279, 127432.

Deo, A. & Chaudhari, S. (2026). Fusion-based movie recommender system combining genre similarity, LightFM, and neural networks. Egyptian Informatics Journal.

Guo, S. et al. (2026). A collaborative process parameter recommender system for fleets of networked manufacturing machines. Journal of Manufacturing Systems, 85, 22–33.

Huang, L. (2022). Design and application of recommendation system model for e-commerce platform based on deep learning algorithm. Proceedings, IEEE.

Ishii, T. et al. (2026). BIG-PU: An evaluation metric for exploration based on preference elicitation in recommender systems. Expert Systems With Applications, 308, 131077.

Jiang, H. (2024). Design of a personalized travel itinerary generation system using deep reinforcement learning and graph neural networks. Proceedings, IEEE.

Li, H. et al. (2021). An intelligent recommendation system for performance equipment operation and maintenance via deep neural network and attention mechanism. IEEE DDCLS 2021.

Li, Y. et al. (2026). STLLMRec: Enhancing explainable recommendation via self-training LLMs. Springer.

Lu, M. et al. (2025). A recommender system for social networks using link prediction, clustering and genetic algorithm. Egyptian Informatics Journal, 32, 100784.

Mall, R. & Singh, M. P. (2026). A novel heterogeneous hypergraph social network recommendation system. User Modeling and User-Adapted Interaction, 36(6). https://doi.org/10.1007/s11257-026-09442-y

Marwan, M. et al. (2026). An ensemble deep learning model based on graph neural networks for intelligent caching in next generation data centers. Discover Computing, 29, 38.

Nikitha, C. & Sushama, C. (2024). A review on deep learning powered online product recommender system in the digital market place. IEEE ISML 2024.

Patil, M. et al. (2026). AI for swimming recommendation systems: exploring the current landscape and research opportunities. Discover Applied Sciences, 8, 156.

Patil, P. R. & Basavaraju, P. H. (2026). A hybrid embedding and graph neural network framework for cross-domain recommendation. Intelligent Network and Systems Society.

Payandenick, M. et al. (2026). A multi-criteria recommendation system for personalised tourism experiences with user query analysis. Information Technology & Tourism, 28(3). https://doi.org/10.1007/s40558-025-00348-w

Peng, B. & Li, D. (2026). Artificial intelligence in digital media, humanities, and information science: a multidimensional analysis of research trends and user perceptions. Springer Nature.

Rajput, I. S. et al. (2025). A PRISMA-based systematic review on deep learning in multi-criteria recommender systems. Procedia Computer Science, 259, 1533–1542.

Ramteke, S. et al. (2025). Enhancing recommendation systems with graph contrastive learning: A comprehensive study. Procedia Computer Science.

Wang, R. et al. (2025). A hybrid recommendation approach integrating matrix decomposition and deep neural networks for enhanced accuracy and generalization. IEEE NNICE 2025.

Wang, Y. & Chen, L. (2025). A decision support system for ideological and political education in universities based on deep neural networks and reinforcement learning. IEEE ICMEIM 2025.

Yang, H. et al. (2022). A review of academic recommendation systems based on intelligent recommendation algorithms. IEEE ICIVC 2022.

Yang, Z. et al. (2026). Dual-path fairness optimization for graph neural network-based recommendation. Scientific Reports (Nature Research).

Yoon, J. & Jang, S. (2023). Evolution of deep learning-based sequential recommender systems: From current trends to new perspectives. Knowledge and Information Systems.

Yu, C. & Aviles, A. (2025). Design and application of recommendation system model for e-commerce platform based on deep learning algorithm. Proceedings, IEEE.

Zangari, A. et al. (2026). Transformers for graph-based recommender systems: A survey. Chinese

Academy of Sciences.

Zarzour, H. et al. (2022). Using deep learning for positive reviews prediction in explainable recommendation systems. Proceedings, IEEE.

Zhang, E. & Wang, D. (2026). Artificial intelligence technology for the ethical issues research from a Marxist perspective under deep learning. Scientific Reports (Nature Research).

Zhai, Q. (2025). Optimization of product recommendation algorithm in social e-commerce using graph neural network. Proceedings, IEEE.

Zou, Y. et al. (2026). Multi-type context-aware conversational recommender system. Information Processing & Management.

Guo, X. et al. (2026). Leveraging hierarchy-aware diffusion model and knowledge-enhanced contrastive learning for recommendation. Springer.

Chen, Z. et al. (2026). Fuzzy-enhanced variable-weight graph convolutional network for recommendation. Computers and Electrical Engineering.

Wang, L. et al. (2026). Graph contrastive learning view construction methods in recommender systems: A survey. Higher Education Press.

Liu, Y. et al. (2026). Group deep neural network approach in semantic recommendation system for movie recommendation in online networks. Springer.

Huang, S. et al. (2026). Hybrid LSTM-SVM with Aquila optimization-based optimal feature selection for effective recommender system. IEICE.

Chen, H. et al. (2026). Knowledge graph-based adaptive recommendation model for training courses. Inderscience Publishers.

Kim, J. et al. (2025). PTDL-Rec: A recommendation model integrating personality traits and deep learning. Neurocomputing.

Zhang, Y. et al. (2026). Performance evaluation of a deep learning based personal recommender system using sentiment analysis and time-series adaptive learning. Proceedings, IEEE.

Li, X. et al. (2026). Personalization in smart urban environments: A taxonomy and survey of recommender systems. Springer.

Wang, M. et al. (2021). Personalized skincare recommender system using deep learning. Proceedings, IEEE.

Chen, Q. et al. (2026). Recommendation system for social platform using deep learning techniques. Proceedings, IEEE.

Sun, Z. et al. (2026). Research on intelligent recommendation system of English teaching resources driven by big data. Proceedings, IEEE.

Wu, L. et al. (2026). Research on optimization of intelligent recommendation system for college employment guidance based on deep learning. Proceedings, IEEE.

Zhang, W. et al. (2026). Research on optimization of personalized algorithm of e-commerce recommendation system based on deep learning. Proceedings, IEEE.

Liu, H. et al. (2026). Research on optimizing personalized recommendation system of brand promotion content with deep learning algorithm. Proceedings, IEEE.

Zhao, L. et al. (2026). Research on optimizing the intelligent recommendation system of college courses by machine learning algorithm. Proceedings, IEEE.

Chen, B. et al. (2026). Research on product recommendation algorithm based on consumer sentiment analysis and neural network. Proceedings, IEEE.

Lin, C. et al. (2025). Research on the methodology of personalized recommendation. Natural Language Processing.

Park, S. et al. (2026). SPECTER-BS: Effective citation recommendation using SPECTER with bibliographic scoring. Springer.

Zhao, Y. et al. (2026). Self-supervised social attentive deep reinforcement learning for recommendation. Engineering Applications of Artificial Intelligence.

Liu, Q. et al. (2024). The deep learning-based physical education course recommendation system. Heliyon.

Wang, Z. et al. (2026). The optimization of vocal music teaching by integrating the STEAM concept with the intelligent recommendation system. Scientific Reports (Nature Research).

Sun, K. et al. (2025). The rising safety concerns of deep recommender systems. The Innovation.

Zhang, X. et al. (2026). Tourism image recognition and scenic spot recommendation system based on deep learning. Proceedings, IEEE.

Tang, R. et al. (2024). Multi-stakeholder recommendation system through deep learning. Information Processing & Management.

Chen, X. et al. (2026). Multi-view debiasing representation learning for recommendation. Information Processing & Management.

Wu, J. et al. (2026). Multimodal deep learning for sports teacher behavior analysis. Scientific Reports (Nature Research).

Li, Z. et al. (2026). Multimodal fusion in graph-based recommendation systems. Springer.

Kim, M. et al. (2026). Consumer psychological modeling and behavior prediction system based on graph neural network. Springer.

Zhao, K. et al. (2021). Content-based clothing recommender system using deep neural network. Proceedings, IEEE.

Sun, Y. et al. (2024). DLSF: Deep learning and semantic fusion based recommendation. Expert Systems with Applications.

Li, M. et al. (2026). Deep collaborative filtering recommender systems in smart cities: A systematic review. Springer.

Chen, Y. et al. (2025). Deep learning-based collaborative filtering recommendation. Procedia Computer Science.

Wang, X. et al. (2026). Deep learning-based modeling of stakeholders preferences. Engineering Applications of Artificial Intelligence.

Liu, X. et al. (2021). Deep learning-based semantic personalization. International Journal of Information Management.

Zhang, Z. et al. (2026). Detecting shilling groups in recommender systems based on user modeling. Neurocomputing.

Chen, L. et al. (2026). Developing and validating a machine learning pharmaceutical therapy recommender system. BioMed Central.

Park, J. et al. (2026). DyHuCoG: A dynamic hypergraph cooperative game for preference-aware recommendation. Intelligent Network and Systems Society.

Liu, W. et al. (2026). E-commerce intelligent recommendation system based on deep learning. Proceedings, IEEE.

Wu, Y. et al. (2026). Efficient incorporation of new interactions in graph recommenders via folding-in. Springer.

Zhang, L. et al. (2026). Enhancing knowledge graph embeddings for improved link prediction using graph neural networks. Springer.

Sun, L. et al. (2026). Enhancing knowledge graph recommendations through deep reinforcement learning. Scientific Reports (Nature Research).

Chen, W. et al. (2026). Enhancing personalized learning experiences by leveraging deep learning for content understanding in e-learning recommender systems. Proceedings, IEEE.

Wang, Q. et al. (2026). Enhancing crop recommendation systems using deep learning techniques on soil and environmental data. Proceedings, IEEE.

Zhao, Z. et al. (2026). Experimental design of personalized recommendation system based on deep learning. Proceedings, IEEE.

Li, W. et al. (2026). Fusion of pretrained model features with style-aware attention for enhanced content-based recommendation system. Springer.

Zhang, R. et al. (2026). Benchmarking heterogeneous network-based methods for drug repurposing. Scientific Reports (Nature Research).

Wang, J. et al. (2021). Adaptable recommendation system for outfit selection with deep learning.

IFAC PapersOnLine, 54(13), 605–610.

Liu, Y. et al. (2025). Building robust deep recommender systems: Utilizing a weighted adversarial noise propagation framework. Proceedings, IEEE.

Sun, H. et al. (2026). CO2-based recommendation system for predictive ventilation scheme. Energy.

Wu, M. et al. (2025). CARES: A hybrid caregivers recommendation system using deep learning. Internet of Things.

Zhang, H. et al. (2026). Unveiling consumer preferences: A theory-based multi-criteria recommender system. International Journal.

Li, Q. et al. (2026). Multimodal deep learning for sports teacher behavior analysis: Design and evaluation of a personalized continuing education recommendation system. Scientific Reports (Nature Research).

Chen, C. et al. (2026). Movie recommendation system with sentiment analysis. Egyptian Informatics Journal.

Wang, C. et al. (2026). Personalization in smart urban environments: A taxonomy and survey of recommender systems. Springer.

Sun, B. et al. (2026). Design of personalized content recommendation system for English reading based on deep learning and intelligent algorithms. Proceedings, IEEE.

Liu, B. et al. (2026). Design of personalized learning resource intelligent recommendation system based on deep learning. Proceedings, IEEE.

Zhang, Q. et al. (2026). Optimization and effect evaluation of deep learning in intelligent advertising recommendation system. Proceedings, IEEE.

Chen, K. et al. (2026). Optimization of personalized recommendation algorithm for e-commerce platforms based on deep learning. Proceedings, IEEE.

College Graduates Employment Recommendation (2026). Using hybrid deep learning based convolutional deep semantic structure modelling. Proceedings, IEEE.