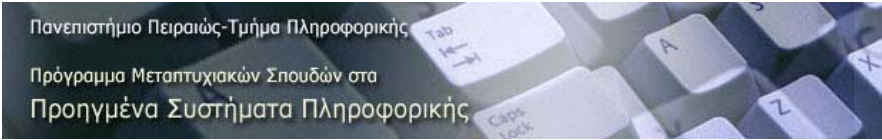




Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής
Πρόγραμμα Μεταπτυχιακών Σπουδών
«Προηγμένα Συστήματα Πληροφορικής»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	Συστήματα Σύστασης βασισμένα σε Τεχνικές Νοημοσύνης Σμήνους
Όνοματεπώνυμο Φοιτητή	Τσεχελίδου Ελένη του Κωνσταντίνου
Αριθμός Μητρώου	ΜΠΣΠ/06046
Κατεύθυνση	Ευφυείς Τεχνολογίες Αλληλεπίδρασης Ανθρώπου-Υπολογιστή
Επιβλέπων	Γεώργιος Α. Τσιχριντζής, Καθηγητής



Πανεπιστήμιο Πειραιώς-Τμήμα Πληροφορικής
Πρόγραμμα Μεταπτυχιακών Σπουδών στα
Προηγμένα Συστήματα Πληροφορικής

Ημερομηνία Παράδοσης **Μάιος 2010**

Τριμελής Εξεταστική Επιτροπή

(υπογραφή)

Γεώργιος Τσιχριντζής
Καθηγητής

(υπογραφή)

Αποστόλου Δημήτριος
Λέκτορας

(υπογραφή)

Κωνσταντόπουλος
Χαράλαμπος
Λέκτορας

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στον επιβλέποντα Καθηγητή κ. Γεώργιο Τσιχριντζή, ο οποίος με τίμησε με την εμπιστοσύνη του και μου έδωσε το ερέθισμα της αναζήτησης της γνώσης σε ένα τόσο ενδιαφέρον επιστημονικό πεδίο.

Επίσης, θα ήθελα να ευχαριστήσω τους φίλους μου, που μου συμπαραστάθηκαν και με στήριξαν σε όλη τη διάρκεια εκπόνησης της μεταπτυχιακής αυτής διατριβής.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΑΙΑ

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

Abstract.....	1
Περίληψη	1
ΕΙΣΑΓΩΓΗ.....	2
ΚΕΦΑΛΑΙΟ 1	4
1.Νοημοσύνη Σμήνους	4
1.1 Βιολογική συλλογική συμπεριφορά	6
1.1.1 Αυτόνομος Πράκτορας	8
1.1.2 Αυτο-οργάνωση	8
1.1.3 Σμήνη και Τεχνητή Ζωή	10
1.2 Particle Swarm Optimization (PSO).....	10
1.3 Ant Colonies Optimization.....	15
1.3.1 Εξόρυξη γνώσης από δεδομένα	15
1.4 Εφαρμογές της Νοημοσύνης Σμήνους στην εξόρυξη γνώσης από δεδομένα	17
1.4.1 Particle Swarm Optimization και εξόρυξη γνώσης από δεδομένα	17
1.4.2 Ant Colony Optimization και εξόρυξη γνώσης από δεδομένα ..	19
ΚΕΦΑΛΑΙΟ 2	21
2.Συστήματα Σύστασης	21
2.1 Σύσταση: Αρχές και παραδείγματα	22
2.1.1 Ένα μοντέλο της διαδικασίας σύστασης.....	22
2.1.2 Θέματα για υπολογιστικά Συστήματα Σύστασης	23
2.2 Τύποι Συστημάτων Σύστασης	24
2.2.1 Συστήματα Σύστασης με βάση το περιεχόμενο	25
2.2.2 Συστήματα Υποστήριξης Συστάσεων.....	26
2.2.3 Social data mining systems.....	27
2.2.4 Συστήματα Συνεργατικής Διοίκησης.....	30
2.3 Διαμόρφωση και στήριξη κοινοτήτων ενδιαφέροντος.....	33
2.3.1 Διαμόρφωση Κοινοτήτων με προστασία του ιδιωτικού απορρήτου	33
2.3.2 Συνδυάζοντας πολλούς τύπους πληροφοριών για τον υπολογισμό συστάσεων	34
2.3.3 Συνδυάζοντας πολλαπλές πηγές προτιμήσεων	35
2.4 Συμπεράσματα	36
ΚΕΦΑΛΑΙΟ 3	37
3. Συστήματα Σύστασης βασισμένα σε Τεχνικές Νοημοσύνης Σμήνους	37
3.1 Εισαγωγή	37
3.2 Recommenders στο e-Commerce	38
3.2.1 Τεχνολογίες της Σύστασης	40
3.3 Recommenders στο e-learning.....	43
3.3.1 Σύστημα πολλών πρακτόρων στην εφαρμογή σύνθετων συστημάτων	45
3.3.2 Biologic Oriented Agentification.....	45
3.4 Συστήματα Σύστασης για e-learning Εφαρμογές.....	46
3.4.1 Αρχιτεκτονική πρόταση	47
3.4.2 Ο αλγόριθμος συσταδοποίησης και ταξινόμησης μηρυγκιών .	48
3.4.3 Συμπεράσματα	52
ΚΕΦΑΛΑΙΟ 4	53
4. Particle Swarm Optimization Recommender System.....	53
4.1 ΕΙΣΑΓΩΓΗ	53
4.2 Βελτιστοποίηση σμήνους σωματιδίων.....	54

4.3 Επισκόπηση του Συστήματος	55
4.3.1 Neighbourhood Selection	56
4.3.2 Ταίριασμα προφίλ.....	56
4.3.3 Best Profile Collection.....	58
4.3.4 Κάνοντας μια σύσταση.....	58
4.3.5 Πειράματα.....	61
4.3.6 Συμπεράσματα	64
ΚΕΦΑΛΑΙΟ 5	65
5.Συμπεράσματα	65
ΑΝΑΦΟΡΕΣ	68
LINKS	76

Abstract

This thesis investigates the application of Swarm Intelligence algorithms to the field of Recommender Systems. The theoretical background of Swarm Intelligence is described with focus on Particle Swarm Optimization and Ant Colony Optimization algorithms. Afterwards, we present the sector of Recommender Systems, which in the past few years had a big growth due to the enormous use of interactive Web sites, which offer numerous choices with regard to services in the fields of e-commerce, e-learning and entertainment. We have described the various types of Recommender Systems, as well as the way that facilitate the individuals to share their opinions and to profit from the experiences of others in the frame of Collaborative Filtering. Finally, we present many algorithms proposed by various researchers.

Key terms: Swarm Intelligence, Self-Organization, Particle Swarm Optimization, Ant Colony Optimization, Data Mining, Web Mining, Collaborative Filtering, Recommender Systems, e-Commerce Recommender Systems, e-Learning Recommender Systems

Περίληψη

Σε αυτήν την εργασία εξετάζουμε, μέσω της πρόσφατης διεθνούς αρθρογραφίας, τα Συστήματα Σύστασης που υλοποιούνται με αλγόριθμους βασισμένους στη Νοημοσύνη Σμήνους. Δίνεται το θεωρητικό υπόβαθρο της Νοημοσύνης Σμήνους και περιγράφονται οι αλγόριθμοι Particle Swarm Optimization και Ant Colony Optimization. Στη συνέχεια, παρουσιάζεται ο τομέας των Συστημάτων Σύστασης, ο οποίος τα τελευταία χρόνια εμφανίζει μεγάλη ανάπτυξη λόγω της τεράστιας διάδοσης αλληλεπιδραστικών δικτυακών τόπων, οι οποίοι προσφέρουν πλήθος επιλογών όσον αφορά σε υπηρεσίες ηλεκτρονικού εμπορίου, ηλεκτρονικής μάθησης και ψυχαγωγίας. Περιγράφονται τα διάφορα είδη Συστημάτων Σύστασης, καθώς και ο τρόπος που επιτρέπουν στα άτομα να μοιράζονται τις απόψεις τους και να επωφελούνται ο ένας από τις εμπειρίες του άλλου στα πλαίσια της Συνεργατικής Διοίκησης. Τέλος, γίνεται μια συνοπτική παρουσίαση υλοποιήσεων από διάφορους ερευνητές.

Λέξεις κλειδιά: Νοημοσύνη Σμήνους, Αυτο-οργάνωση, Βελτιστοποίηση Σμήνους Σωματιδίων, Βελτιστοποίηση Αποικίας Μυρμηγκιών, Εξόρυξη Γνώσης από Δεδομένα, Εξόρυξη Γνώσης από το Διαδίκτυο, Συνεργατική Διοίκηση, Συστήματα Συστάσεων, Συστήματα Συστάσεων Ηλεκτρονικού Εμπορίου, Συστήματα Συστάσεων Ηλεκτρονικής Μάθησης.

ΕΙΣΑΓΩΓΗ

Το διαδίκτυο και ο παγκόσμιος Ιστός δημιούργησαν έναν κόσμο απεριόριστων δυνατοτήτων: οι χρήστες του διαδικτύου έρχονται καθημερινά αντιμέτωποι με αλληλεπιδραστικούς τόπους Web, μπορούν να ακούσουν μουσική, να συμμετάσχουν σε συνομιλίες, και μπορούν να παραγγείλουν οποιοδήποτε καταναλωτικό προϊόν. Αλλά αυτός ο κόσμος προσφέρει ατελείωτες επιλογές: Πώς να επιλέξει κανείς από ένα τεράστιο σύμπαν προϊόντων και υπηρεσιών διαφορετικής ποιότητας;

Για την αντιμετώπιση αυτού του ζητήματος προέκυψαν υπολογιστικά Συστήματα Συστάσεων (Recommender Systems), τα οποία επιτρέπουν στα άτομα να μοιράζονται τις απόψεις τους και να επωφελούνται ο ένας από τις εμπειρίες του άλλου. Η εργασία αυτή παρουσιάζει το πλαίσιο για την κατανόηση των Συστημάτων Σύστασης και ερευνά μια σειρά προσεγγίσεων σχετικά με αυτό το πλαίσιο. Δύο είναι οι κύριες προκλήσεις που αντιμετωπίζει η έρευνα:

- (1) η βοήθεια προς τα άτομα, ώστε να σχηματίσουν κοινότητες ενδιαφέροντος σεβόμενοι την ιδιωτικότητά τους και
- (2) η ανάπτυξη αλγορίθμων που να συνδυάζουν πολλούς τύπους πληροφορίας για τον υπολογισμό των συστάσεων.

Η αποτελεσματική πρόσβαση στην ηλεκτρονική πληροφόρηση είναι ένα πρόβλημα που συγκεντρώνει πολλά θέματα τεχνητής νοημοσύνης: όπως η αναπαράσταση της γνώσης, η φυσική γλώσσα, η μοντελοποίηση του χρήστη, η μηχανική μάθηση, καθώς και distributed computing, Human-Computer Interaction και βάσεις δεδομένων. Τα τελευταία χρόνια, οι τεχνικές Τεχνητής Νοημοσύνης έχουν παράσχει ουσιαστική πρόοδο προς την επίλυση των προβλημάτων πρόσβασης πληροφοριών μέσα από τους ακόλουθους τρεις τομείς: υποδείγματα και τεχνικές για Συστήματα multi-agents, ανάκτηση πληροφοριών και Συστήματα Σύστασης. Στον πρώτο τομέα, υπάρχει μια μεγάλη Κοινότητα που αναπτύσσει αρχιτεκτονικές, γλώσσες, μοντέλα και τεχνικές για την αλληλεπίδραση και την επικοινωνία μεταξύ των πρακτόρων. Στον τομέα της ανάκτησης πληροφοριών έχουν αναπτυχθεί τεχνικές για την αντιμετώπιση του ετερογενούς χώρου και των δυναμικών πληροφοριών που αντιπροσωπεύονται από το World Wide Web. Ωστόσο, τα συστήματα ανάκτησης πληροφοριών παραδοσιακά δεν ενδιαφέρονται για τους επιμέρους χρήστες, προτιμώντας να επικεντρωθούν στην ανάπτυξη γενικών τεχνικών και όχι σε τεχνικές προσαρμοσμένες στις ανάγκες μεμονωμένων ατόμων. Αυτό το κενό έρχονται να συμπληρώσουν τα Συστήματα Σύστασης, τα οποία επικεντρώνονται στην αλληλεπίδραση ανάμεσα στην ανάκτηση της πληροφορίας και στο μοντέλο του χρήστη, προκειμένου να παράσχουν μια πιο εξατομικευμένη εμπειρία ανάκτησης και να βοηθήσουν τους χρήστες να επιλέξουν μεταξύ εναλλακτικών λύσεων ανάκτησης και να βελτιώσουν τα ερωτήματά τους.

Τα θέματα τα οποία τίγονται είναι η απόκτηση των προφίλ των χρηστών βασισμένη στα δημογραφικά δεδομένα, στην ιστορικότητα, στις προτιμήσεις του χρήστη, στις πολιτισμικές και στις κοινωνικές πλευρές, η ανάκτηση εξατομικευμένης πληροφορίας, το απόρρητο και η ασφάλεια των δεδομένων του χρήστη. Η επικοινωνία μεταξύ των πρακτόρων (διαπραγμάτευση, συντονισμός, ανταλλαγή γνώσης μεταξύ των πρακτόρων), η συνεργατική διοίκηση, το φιλτράρισμα του περιεχομένου. Τα πεδία εφαρμογής περιλαμβάνουν, αλλά δεν περιορίζονται στα e-Commerce, e-Learning, Intelligent Tutoring Systems, Web information retrieval, Web services.

Στην παρούσα εργασία διερευνήσαμε υλοποιήσεις Συστημάτων Σύστασης βασισμένες σε αλγόριθμους βασισμένους στη Νοημοσύνη Σμήνους. Περιγράφηκε το θεωρητικό υπόβαθρο της Νοημοσύνης Σμήνους και πιο συγκεκριμένα οι αλγόριθμοι Particle Swarm Optimization και Ant Colony Optimization. Ακόμη, έγινε εκτενής αναφορά στις διάφορες κατηγορίες των Συστημάτων Σύστασης: τα content-based Συστήματα Σύστασης, τα Συστήματα Υποστήριξης Σύστασης (Recommendation support systems), τα Συστήματα Εξόρυξης Κοινωνικών Δεδομένων (Social data mining systems) και τα Συστήματα Συνεργατικής Διοίκησης (Collaborative filtering systems).

Συγκεκριμένα, τα content-based Συστήματα Σύστασης (συστήματα σύστασης με βάση το περιεχόμενο) είναι συστήματα που χρησιμοποιούν μόνο τις προτιμήσεις του χρήστη και προτείνουν προϊόντα που είναι παρόμοια με αυτά που άρεσαν στο χρήστη στο παρελθόν. Επικεντρώνονται σε αλγόριθμους για μάθηση των προτιμήσεων του χρήστη και στο φιλτράρισμα νέων προϊόντων που να ταιριάζουν περισσότερο με τις προτιμήσεις του χρήστη.

Τα Συστήματα Υποστήριξης Σύστασης (Recommendation support systems) δεν αυτοματοποιούν τη διαδικασία σύστασης, ως εκ τούτου, δεν πρέπει να παρουσιάσουν προτιμήσεις ή να υπολογίσουν συστάσεις. Αντίθετα, χρησιμεύουν ως εργαλεία για την υποστήριξη ατόμων στην κοινή χρήση συστάσεων, βοηθώντας τόσο αυτούς που παράγουν συστάσεις, όσο και αυτούς που αναζητούν μια σύσταση.

Τα Συστήματα Εξόρυξης Κοινωνικών Δεδομένων (Social data mining systems) εξορύσσουν τις έμμεσες προτιμήσεις από υπολογιστικές εγγραφές κοινωνικής δραστηριότητας, όπως μηνύματα Usenet, το ιστορικό χρήσης ενός συστήματος, αναφορές ή υπερ-συνδέσεις. Αυτά τα συστήματα επίσης επικεντρώνονται στα ζητήματα HCI (Human computer Interaction), που αφορούν σε οπτικοποίηση των αποτελεσμάτων. Αυτές οι παρουσιάσεις συχνά έχουν χρησιμοποιηθεί για να βοηθήσουν την πλοήγηση των διαστημάτων πληροφορίας όπως το World Wide Web και αυτό βοήθησε στην καθιέρωση του όρου social navigation (κοινωνική πλοήγηση).

Τα Συστήματα Συνεργατικής διοίκησης (Collaborative filtering systems) απαιτούν από τους χρήστες να εκφράσουν προτιμήσεις αξιολογώντας μια δωδεκάδα ή δύο αντικείμενα, συγχωνεύοντας τους ρόλους αυτού που ζητάει τη σύσταση και αυτού που παρέχει την προτίμηση. Αυτά τα συστήματα επικεντρώνονται σε αλγόριθμους για το ταίριασμα ατόμων με βάση τις προτιμήσεις τους και τη στάθμιση των ενδιαφερόντων των ατόμων με παρόμοια γούστα για να παράξουν μια σύσταση για τον χρήστη που τη ζητάει.

Τέλος, γίνεται εκτενής αναφορά για τις υλοποιήσεις που υπάρχουν στην επιστημονική κοινότητα σχετικά με τα Συστήματα Σύστασης Ηλεκτρονικού Εμπορίου (e-Commerce Recommender Systems) και τα Συστήματα Σύστασης Ηλεκτρονικής Μάθησης (e-Learning Recommender Systems).

ΚΕΦΑΛΑΙΟ 1

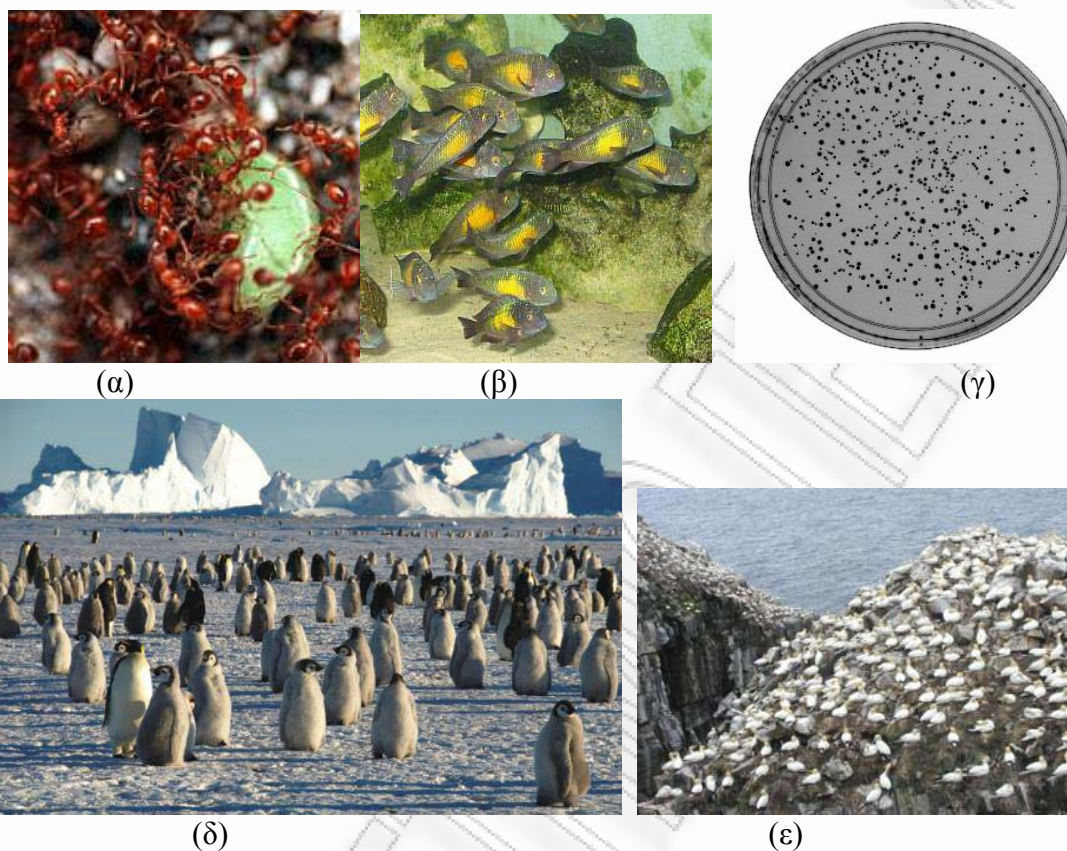
1.Νοημοσύνη Σμήνους

Η Νοημοσύνη Σμήνους (**Swarm intelligence (SI)**) αποτελεί έναν διεπιστημονικό τομέα με ευρύ πεδίο εφαρμογών. Οι εξελίξεις στην κατανόηση και την αξιοποίηση των συστημάτων που βασίζονται στις αρχές της Νοημοσύνης Σμήνους περιλαμβάνει θέματα όπως είναι η μοντελοποίηση και η ανάλυση των συλλογικών βιολογικών συστημάτων, η εφαρμογή των βιολογικών μοντέλων της Νοημοσύνης Σμήνους σε πραγματικά προβλήματα, η θεωρητική και εμπειρική έρευνα στη βελτιστοποίηση αποικίας μυρμηγκιών (Ant Colony Optimization), η βελτιστοποίηση σμήνους σωματιδίων (Particle Swarm Optimization), η ρομποτική σμήνους (swarm robotics) και άλλοι αλγόριθμοι Νοημοσύνης Σμήνους.

Η Νοημοσύνη Σμήνους είναι μια τεχνική Τεχνητής Νοημοσύνης βασισμένη στη μελέτη της συλλογικής συμπεριφοράς σε αποκεντρωμένα, αυτο-οργανωμένα συστήματα. Ο όρος "Swarm Intelligence" εισήχθει από τους Beni & Wang το 1989 στο πλαίσιο κυτταρικών ρομποτικών συστημάτων. Τα SI συστήματα αποτελούνται συνήθως από έναν πληθυσμό απλών πρακτόρων (agents) ή boids (birds-ooids) που αλληλεπιδρούν τοπικά μεταξύ τους και με το περιβάλλον τους. Οι πράκτορες ακολουθούν πολύ απλούς κανόνες, και παρόλο που δεν υπάρχει κεντρική δομή ελέγχου που να υπαγορεύει πώς θα πρέπει να συμπεριφέρονται οι επιμέρους πράκτορες, τοπικά και σε κάποιο βαθμό τυχαία, οι αλληλεπιδράσεις μεταξύ αυτών των πρακτόρων οδηγούν στην εμφάνιση "ευφυούς" συνολικής συμπεριφοράς, άγνωστης στους επιμέρους παράγοντες. Φυσικά παραδείγματα SI περιλαμβάνουν τις αποικίες των μυρμηγκιών, σμήνη πουλιών, αγέλες ζώων, και κοπάδια ψαριών.

Η Νοημοσύνη Σμήνους είναι μια ιδιότητα συστημάτων μη ευφυών πρακτόρων που επιδεικνύουν συλλογικά ευφυή συμπεριφορά. Ένας πράκτορας σε αυτόν τον ορισμό αντιπροσωπεύει μια οντότητα ικανή να ανιχνεύσει το περιβάλλον του και να επιχειρήσει απλή επεξεργασία περιβαλλοντικών παρατηρήσεων, προκειμένου να εκτελέσει μια ενέργεια που έχει επιλέξει ανάμεσα στις διαθέσιμες σε αυτόν. Αυτές οι ενέργειες περιλαμβάνουν την τροποποίηση του περιβάλλοντος στο οποίο λειτουργεί ο πράκτορας. Η επικοινωνία μεταξύ των κοινωνικών εντόμων επιτυγχάνεται έμμεσα με μια μέθοδο την οποία ο εντομολόγος Grassé έχει ονομάσει "stigmergy". Ευφυής συμπεριφορά συχνά προκύπτει μέσω έμμεσης επικοινωνίας μεταξύ των πρακτόρων, και αυτή είναι η αρχή της "stigmergy". Θα πρέπει να τονιστεί, ωστόσο, ότι οι επιμέρους πράκτορες δεν έχουν ρητή γνώση επίλυσης προβλημάτων και η ευφυής συμπεριφορά προκύπτει ως αποτέλεσμα των δράσεων των κοινωνιών των εν λόγω πρακτόρων.

Παρόλο που κατά κανόνα δεν υπάρχει κεντρική δομή ελέγχου που να υπαγορεύει τη συμπεριφορά των επιμέρους πρακτόρων, οι τοπικές αλληλεπιδράσεις μεταξύ αυτών των πρακτόρων συχνά οδηγεί στην εμφάνιση μιάς γενικής συμπεριφοράς. Τα μεμονωμένα μυρμηγκία είναι συμπεριφορικά απλά έντομα με περιορισμένη μνήμη και επιδεικνύουν δραστηριότητα που έχει ένα στοιχείο τυχαιότητας. Παραδείγματα συστημάτων όπως αυτό μπορούν να βρεθούν στη φύση, και συμπεριλαμβάνουν αποικίες μυρμηγκιών, σμήνη πουλιών, αγέλες ζώων, καλλιέργειες βακτηρίων, και κοπάδια ψαριών (Σχήμα 1). Πρόσφατα, η μεταφορά των κοινωνικών εντόμων για την επίλυση προβλημάτων έχει προσελκύσει μεγάλη προσοχή από πολλά διαφορετικά πεδία. Σε αυτήν την ενότητα, θα εισαγάγουμε δύο από τους πιο δημοφιλείς αλγορίθμους βελτιστοποίησης της ευφύιας σμήνους: το σύστημα αποικίας μυρμηγκιών (Ant Colony System) και τη βελτιστοποίηση σμήνους σωματιδίων (Particle Swarm Optimization - PSO).



Σχήμα 1: Μερικά παραδείγματα αποικιών από τη φύση: (α) αποικία μυρμηγκιών, (β) αποικία ψαριών, (γ) αποικία βακτηριδίων, (δ) αποικία πγκουίνων, (ε) αποικία πουλιών

Η Νοημοσύνη Σμήνους (Swarm Intelligence (SI)) είναι ένα καινοτόμο παράδειγμα καταναμημένης ευφυίας για την επίλυση προβλημάτων βελτιστοποίησης που αρχικά εμπνεύστηκε από τα βιολογικά παραδείγματα των φαινομένων δημιουργίας σμηνών, κοπαδιών και αγελών στα σπονδυλωτά. Η Βελτιστοποίηση Σμήνους Σωματιδίων (Particle Swarm Optimization - PSO) ενσωματώνει συμπεριφορές που παρατηρούνται στα σμήνη των πτηνών, των μελισσών και των μυρμηγκιών, καθώς και των κοπαδιών των ψαριών, ακόμη και ανθρώπινες κοινωνικές συμπεριφορές, από τις οποίες και προέκυψε η ιδέα. Η Βελτιστοποίηση Αποικίας Μυρμηγκιών (Ant Colony Optimization - ACO) ασχολείται με τεχνητά συστήματα που είναι εμπνευσμένα από τη συμπεριφορά αναζήτησης τροφής των πραγματικών μυρμηγκιών, τα οποία χρησιμοποιούνται για την επίλυση διακριτών προβλημάτων βελτιστοποίησης.

Ιστορικά, η έννοια της εξεύρεσης χρήσιμων προτύπων στα δεδομένα εμφανίζεται με μια ποικιλία ονομάτων, όπως εξόρυξη δεδομένων, ανακάλυψη γνώσης, εξόρυξη πληροφοριών κ.λπ. Η Εξόρυξη Δεδομένων είναι μια αναλυτική διαδικασία που έχει σχεδιαστεί για να εξερευνήσει μεγάλες ποσότητες δεδομένων σε αναζήτηση συνεπών προτύπων ή/και συστηματικές σχέσεις μεταξύ των μεταβλητών, και στη συνέχεια να επικυρώσει τα συμπεράσματα, εφαρμόζοντας τα πρότυπα που εντοπίστηκαν σε νέα υποσύνολα των δεδομένων. Προκειμένου να επιτευχθεί αυτό, η εξόρυξη δεδομένων

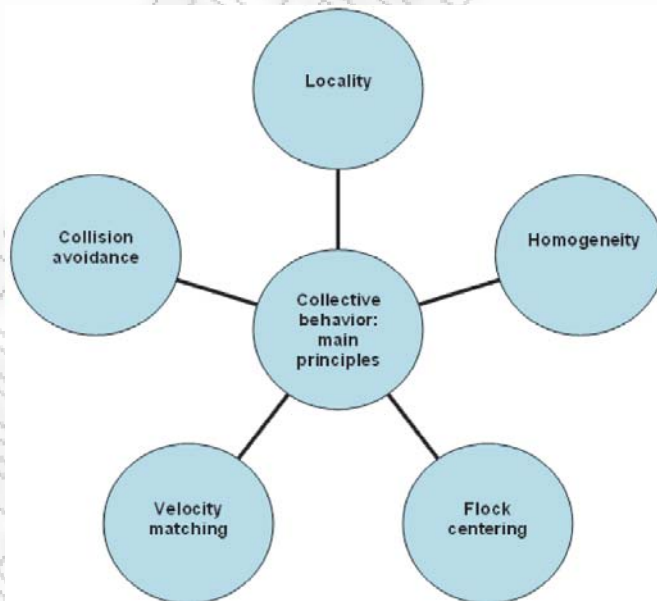
χρησιμοποιεί υπολογιστικές τεχνικές από την στατιστική, την μηχανική μάθηση και την αναγνώριση προτύπων.

Η Εξόρυξη γνώσης από δεδομένα και η ευφυία σμήνους μπορεί να φαίνεται ότι δεν έχουν πολλές κοινές ιδιότητες. Ωστόσο, πρόσφατες μελέτες δείχνουν ότι μπορούν να χρησιμοποιηθούν μαζί για διάφορα προβλήματα εξόρυξης δεδομένων στον πραγματικό κόσμο, ειδικά όταν άλλες μέθοδοι θα ήταν πολύ ακριβές ή δύσκολο να εφαρμοστούν.

1.1 Βιολογική συλλογική συμπεριφορά

Τη συμπεριφορά σμήνους τη βλέπουμε στα σμήνη των πουλιών, στα κοπάδια των ψαριών, καθώς επίσης και σε έντομα όπως τα κουνούπια. Πολλές ομάδες ζώων όπως τα κοπάδια των ψαριών και τα σμήνη των πουλιών επιδεικνύουν σαφή δομική τάξη, με τη συμπεριφορά των οργανισμών τόσο ενσωματωμένη, έτσι ώστε ακόμα κι αν αλλάξουν σχήμα και κατεύθυνση, εμφανίζονται να κινούνται σαν μια ενιαία συνεκτική οντότητα (Couzin et al 2002). Οι κύριες αρχές της συλλογικής συμπεριφοράς όπως παρουσιάζεται στο Σχήμα 2 είναι:

- Ομοιογένεια (Homogeneity): κάθε πουλί στο σμήνος έχει το ίδιο πρότυπο συμπεριφοράς. Το σμήνος κινείται χωρίς έναν ηγέτη, ακόμα κι αν εμφανίζονται προσωρινοί ηγέτες.
- Θέση (Locality): η κίνηση κάθε πουλιού επηρεάζεται μόνο από τους κοντινότερους συντρόφους του στο σμήνος. Η όραση θεωρείται η σημαντικότερη αίσθηση για την οργάνωση του σμήνους.
- Αποφυγή σύγκρουσης (Collision Avoidance): αποφυγή των κοντινότερων συντρόφων του σμήνους.
- Ταίριασμα ταχύτητας (Velocity Matching): προσπάθεια να ταιριαστεί η ταχύτητα μ' αυτή των κοντινών συντρόφων.
- Στροφή του σμήνους (Flock Centering): προσπάθεια του κάθε μέλους να μείνει κοντά στους κοντινούς του συντρόφους.



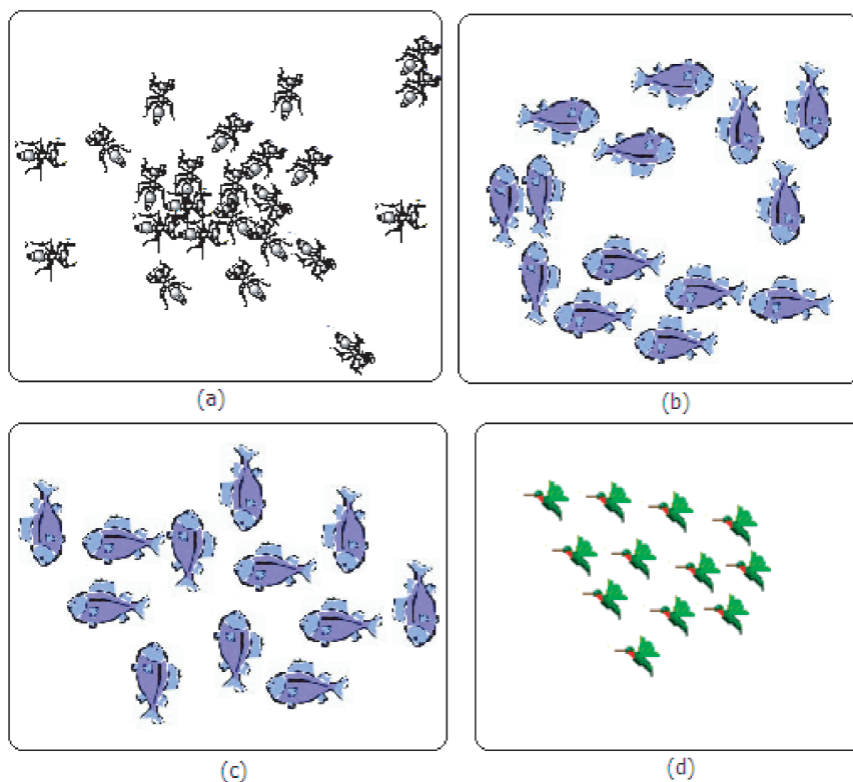
Σχήμα 2: Οι κύριες αρχές της συλλογικής συμπεριφοράς

Τα άτομα πάντα προσπαθούν να διατηρήσουν μια ελάχιστη απόσταση μεταξύ των ιδίων και των άλλων. Αυτός ο κανόνας έχει την πιο υψηλή προτεραιότητα και αντιστοιχεί σε μια συχνά παρατηρηθείσα

συμπεριφορά των ζώων στη φύση. Εάν τα άτομα δεν εκτελούν έναν ελιγμό αποφυγής, τείνουν να προσελκύνονται προς άλλα άτομα (για να αποφύγουν να απομονωθούν) και να ευθυγραμμίζονται με τα γειτονικά τους άτομα.

Οι Couzin et al. προσδιόρισαν τέσσερις συλλογικές δυναμικές συμπεριφορές όπως φαίνονται στο Σχήμα 3:

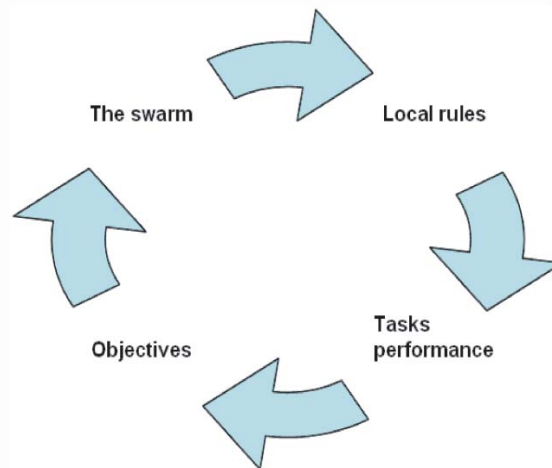
- Σμήνος (Swarm): ένα σύνολο με συνοχή, αλλά με χαμηλό επίπεδο πόλωσης (παράλληλη ευθυγράμμιση) μεταξύ των μελών του
- Δακτύλιος (Torus): τα άτομα περιστρέφονται ασταμάτητα γύρω από έναν κενό πυρήνα (άλεση). Η κατεύθυνση της περιστροφής είναι τυχαία.
- Δυναμική παράλληλη ομάδα (Dynamic parallel group): τα άτομα είναι πολωμένα και κινούνται σαν μια συνεκτική ομάδα, αλλά τα άτομα μπορούν να κινηθούν διαμέσου της ομάδας και η πυκνότητα και η μορφή της ομάδας μπορούν να διακυμανθούν.
- Ιδιαίτερα παράλληλη ομάδα (Highly parallel group): περισσότερο στατική από άποψη ανταλλαγής των χωρικών θέσεων μέσα στην ομάδα απ' ό,τι η δυναμική παράλληλη ομάδα. Η διακύμανση στην πυκνότητα και στη μορφή είναι ελάχιστη.



Σχήμα 3: Διάφορα μοντέλα συλλογικής συμπεριφοράς: (α) σμήνος, (β) δακτύλιος, (γ) δυναμική παράλληλη ομάδα και (δ) ιδιαίτερα παράλληλη ομάδα

Σε υψηλό επίπεδο, ένα σμήνος μπορεί να θεωρηθεί σαν ομάδα πρακτόρων που συνεργάζονται για να πετύχουν κάποιο στόχο (Σχήμα 4). Αυτή η συλλογική νοημοσύνη φαίνεται να προκύπτει συχνά από μεγάλες ομάδες σχετικά απλών πρακτόρων. Οι πράκτορες χρησιμοποιούν απλούς τοπικούς κανόνες Συστήματα Σύστασης βασισμένα σε Τεχνικές Νοημοσύνης Σμήνους

για να κουμαντάρουν τις ενέργειές τους και μέσω των αλληλεπιδράσεων ολόκληρης της ομάδας, το σμήνος επιτυγχάνει τους στόχους του.



Σχήμα. 4. Το απλό σχήμα ενός σμήνους.

1.1.1 Αυτόνομος Πράκτορας

Ένας αυτόνομος πράκτορας (Autonomous Agent) είναι ένα υποσύστημα που αλληλεπιδρά με το περιβάλλον του, το οποίο πιθανώς αποτελείται από άλλους πράκτορες, αλλά ενεργεί σχετικά ανεξάρτητα από όλους τους άλλους πράκτορες. Ο αυτόνομος πράκτορας δεν ακολουθεί εντολές από έναν ηγέτη, ή κάποιο γενικό σχέδιο. Παραδείγματος χάριν, για να συμμετάσχει ένα πουλί σε ένα σμήνος, ρυθμίζει απλά τις κινήσεις του ώστε να συνταχθεί με τις κινήσεις των συντρόφων του στο σμήνος, δηλαδή των γειτόνων του, που είναι κοντά του στο σμήνος. Ένα πουλί σε ένα σμήνος προσπαθεί απλά να μείνει κοντά στους γείτονές του, αλλά να αποφύγει τις συγκρούσεις μαζί τους. Κάθε πουλί δεν παίρνει εντολές από οποιοδήποτε πουλί αρχηγό από τη στιγμή που δεν υπάρχει κανένα πουλί αρχηγός. Οποιοδήποτε πουλί μπορεί να είναι στο μέτωπο, στο κέντρο ή στο πίσω μέρος του σμήνους. Η συμπεριφορά του σμήνους βοηθά τα πουλιά να εκμεταλλευτούν διάφορες καταστάσεις συμπεριλαμβανομένης της προστασίας από τα αρπακτικά ζώα (ειδικά για τα πουλιά στη μέση του σμήνους), και της αναζήτησης τροφής (ουσιαστικά κάθε πουλί εκμεταλλεύεται τα μάτια κάθε άλλου πουλιού).

Ένας τύπος αυτο-οργάνωσης προκύπτει από τη συλλογή των ενεργειών της ομάδας. Εδώ, θα ήταν σκόπιμο να ορίσουμε την έννοια της αυτοοργάνωσης.

1.1.2 Αυτο-οργάνωση

Ένα αυτο-οργανωμένο σύστημα (Self-organizing) όχι μόνο ρυθμίζει ή προσαρμόζει τη συμπεριφορά του, αλλά δημιουργεί τη δική του οργάνωση. Ορίζουμε την οργάνωση ως δομή με λειτουργικότητα. Δομή σημαίνει ότι τα στοιχεία ενός συστήματος είναι ταξινομημένα με συγκεκριμένη σειρά. Αυτό απαιτεί και συνδέσεις που ολοκληρώνουν τα τμήματα σε ένα σύνολο, και διαχωρισμούς που διαφοροποιούν τα υποσυστήματα, έτσι ώστε να αποφεύγονται οι παρεμβολές. Λειτουργικότητα σημαίνει ότι η δομή αυτή πληροί ένα σκοπό (F. Heylighen & C. Gershenson, 2003).

Αυτο-οργάνωση (Self-Organization) λοιπόν, σημαίνει ότι μια λειτουργική δομή εμφανίζεται και διατηρείται αυθόρμητα. Ο έλεγχος που απαιτείται για την επίτευξη του στόχου αυτού, πρέπει να είναι

καταμετρημένοι σε όλα τα συμμετέχοντα στοιχεία. Εάν αυτός ήταν συγκεντρωμένος σε ένα υποσύστημα ή λειτουργική μονάδα, τότε αυτό το υποσύστημα θα μπορούσε να καταργηθεί και το σύστημα θα έχανε την οργάνωσή του. Εάν βγάλουμε το chip του επεξεργαστή από έναν υπολογιστή, αυτός καθίσταται άχρηστος. Εάν πάρουμε οποιοδήποτε μικρό κομμάτι ιστού από έναν ζωντανό εγκέφαλο (όπως συνήθως συμβαίνει κατά τη διάρκεια της χειρουργικής εγκέφαλου), ο εγκέφαλος θα συνεχίσει να λειτουργεί λιγότερο ή περισσότερο, όπως πριν.

Τα αυτο-οργανωμένα συστήματα είναι εγγενώς ισχυρά: μπορούν να αντέξουν μια ποικιλία σφαλμάτων ή ακόμη και μερική καταστροφή. Θα επιδιορθώσουν τις περισσότερες ζημιές τα ίδια, επιστρέφοντας στην αρχική τους κατάσταση. Όταν η ζημία είναι πολύ μεγάλη, η λειτουργία τους θα αρχίσει να επιδεινώνεται, χωρίς ξαφνική κατάρρευση θα αναπροσαρμόσουν την οργάνωσή τους σε οποιοδήποτε αλλαγές στο περιβάλλον, μαθαίνοντας νέα "κόλπα" για την αντιμετώπιση απρόβλεπτων προβλημάτων. Από το χάος δημιουργούν τάξη. Φαινομενικά τυχαίες αναταράξεις βοηθούν και όχι εμποδίζουν την επίτευξη μιας όλο και καλύτερης οργάνωσης.

Ένα αυτο-οργανωμένο σύστημα αποτελείται από έναν μεγάλο αριθμό αντιδρώντων στοιχείων όπως μόρια, νευρώνες, έντομα ή άτομα. Το σύστημα είναι δυναμικό: τα στοιχεία αλλάζουν συνεχώς κατάσταση σε σχέση με τα άλλα. Αλλά εξαιτίας της αμοιβαίας εξάρτησης οι αλλαγές δεν είναι αυθαίρετες: κάποιες σχετικές καταστάσεις είναι "προτιμητέες", υπό την έννοια ότι θα ενισχυθούν ή θα σταθεροποιηθούν, ενώ άλλες θα παρεμποδιστούν ή θα εξαλειφθούν.

Οι αλλαγές αρχικά είναι τοπικές (local): τα στοιχεία αλληλεπιδρούν μόνο με τους άμεσους "γείτονές" τους. Αυτά είναι σχεδόν ανεξάρτητα από στοιχεία που βρίσκονται πιο μακριά. Αλλά η αυτοοργάνωση συχνά ορίζεται ως καθολική (global) τάξη που αναδύεται από τοπικές αλληλεπιδράσεις.

Ένα γνωστό παράδειγμα αυτο-οργάνωσης είναι ο τρόπος με τον οποίο τα μυρμηγκία αφήνουν μονοπάτια φερομόνης μεταξύ των διαφόρων πηγών τροφής. Αρχικά, τα μυρμηγκία εξερευνούν και αφήνουν φερομόνες τυχαία, αλλά οι "καλές" διαδρομές, που οδηγούν σε πλούσιες πηγές μέσω γρήγορων διαδρομών ενισχύονται με θετική ανάδραση, ενώ οι «κακές διαδρομές» εξατμίζονται τελικά. Έτσι, οι πηγές τροφής οργανώνονται σε ένα πυκνό και αποτελεσματικό δίκτυο μονοπατιών αναζήτησης τροφής.

Ο Marco Dorigo, πρωτοπόρος στον τομέα των "αλγόριθμων μυρμηγκιών", έδειξε πώς ένας παρόμοιος μηχανισμός μπορεί να αντιμετωπίσει μια ποικιλία υπολογιστικών προβλημάτων, όπως το περίφημο πρόβλημα του περιοδεύοντος πωλητή (Bonabeau et al., 1998). Μια σημαντική εφαρμογή είναι η δρομολόγηση των μηνυμάτων κατά μήκος των κόμβων και των συνδέσεων του δικτύου επικοινωνίας: οι πιο αποτελεσματικές διαδρομές ενισχύονται, οι λιγότερο αποτελεσματικές εγκαταλείπονται. Η οργάνωση που προκύπτει προσαρμόζεται σε πραγματικό χρόνο: Εάν οι διαδρομές συμφορηθούν, η προτεραιότητά τους υποβαθμίζεται άμεσα και εξετάζονται νέες διαδρομές.

Μια παρόμοια ιδέα έχει εφαρμοστεί για την οργάνωση των υπερ-συνδέσεων στον παγκόσμιο ιστό (Bollen & Heylighen, 1996, Heylighen & Bollen, 2002). Οι web αλγόριθμοι εκμάθησης ενισχύουν τα μονοπάτια των συνδέσεων που οι χρήστες χρησιμοποιούν συχνά, αντικαθιστώντας τα τελικά από μια μόνο σύνδεση. Ενώ οι χρήστες αποφασίζουν τοπικά ποιά σύνδεση να εξερευνήσουν στη συνέχεια, τα αποτελέσματα της επιλογής τους διαδίδονται σε όλο το web. Αυτό θα πρέπει να οδηγήσει τελικά σε μια παγκόσμια τάξη: ιστοσελίδες που «ταιριάζουν» μεταξύ τους με την έννοια ότι η μια είναι πολύ σχετική για τους χρήστες της άλλης, συνδέονται άμεσα. Με αποτέλεσμα τα έγγραφα που αφορούν το ίδιο θέμα να συγκεντρώνονται μαζί. Αυτό το σύνολο εγγράφων μπορεί τώρα να αντιπροσωπεύεται από ένα έγγραφο υψηλότερου επιπέδου, το οποίο αποτελεί ευρετήριο ή περίληψη για το θέμα. Υψηλότερου επιπέδου στοιχεία συνδέονται σε κατηγορίες και αυτές περαιτέρω σε υπερκατηγορίες. Τελικά, ο ιστός σαν σύνολο μπορεί να αυτοοργανώνεται σε ένα αποτελεσματικό, ιεραρχικά διαρθρωμένο δίκτυο συνάψεων, που προσαρμόζεται συνεχώς σε νεοεισαγόμενα έγγραφα, σε αλλαγές στις απαιτήσεις των χρηστών, κ.λ.π..

1.1.3 Σμήνη και Τεχνητή Ζωή

Από το 1990 έχουν προταθεί αρκετοί αλγόριθμοι εμπνευσμένοι από συλλογικές συμπεριφορές (όπως τα κοινωνικά έντομα και τα σμήνη πουλιών). Οι τομείς εφαρμογής αυτών των αλγορίθμων αναφέρονται σε καλά μελετημένα προβλήματα βελτιστοποίησης όπως τα NP-hard (non-deterministic polynomial-time hard) προβλήματα (Traveling Salesman Problem, Quadratic Assignment Problem, Graph problems), τα προβλήματα δρομολόγησης δικτύων, ομαδοποίησης, εξόρυξης γνώσης από δεδομένα, προγραμματισμού εργασιών κ.λπ.

Οι αλγόριθμοι Particle Swarm Optimization (PSO) (βελτιστοποίηση σμήνους σωματιδίων) και Ant Colonies Optimization (ACO) (βελτιστοποίηση αποικιών μυρμηγκιών) είναι αυτήν την περίοδο οι δημοφιλέστεροι αλγόριθμοι στο πεδίο της Νοημοσύνης Σμήνους.

1.2 Particle Swarm Optimization (PSO)

Ο Particle Swarm Optimization αλγόριθμος είναι ένας αλγόριθμος αναζήτησης βασισμένος στον πληθυσμό (population-based search algorithm) και αρχικοποιείται με έναν πληθυσμό τυχαίων λύσεων, που τις αποκαλούμε σωματίδια (particles). Αντίθετα από τις άλλες εξελικτικές τεχνικές υπολογισμών, κάθε σωματίδιο στον PSO συνδέεται επίσης με μια ταχύτητα. Τα σωματίδια κινούνται διαμέσου του χώρου αναζήτησης με ταχύτητες που ρυθμίζονται δυναμικά σύμφωνα με τις ιστορικές συμπεριφορές τους. Επομένως, τα σωματίδια έχουν την τάση να κινηθούν προς την καλύτερη περιοχή αναζήτησης κατά τη διάρκεια της διαδικασίας αναζήτησης. Ο PSO αρχικά σχεδιάστηκε για να μιμηθεί τα πουλιά που αναζητούν τροφή, το οποίο ορίζεται ως ένα «cornfield διάνυσμα».

Η βελτιστοποίηση σμήνους σωματιδίων (PSO) είναι ένας γενικός αλγόριθμος βελτιστοποίησης για την αντιμετώπιση προβλημάτων στα οποία μια καλύτερη λύση μπορεί να αντιπροσωπευτεί ως ένα σημείο ή μια επιφάνεια σε ένα χώρο n-διαστάσεων. Οι υποθέσεις σχεδιάζονται σε αυτόν το χώρο και τους δίνεται μια αρχική ταχύτητα, σαν ένα κανάλι επικοινωνίας μεταξύ των σωματιδίων. Τα σωματίδια στη συνέχεια μετακινούνται στον χώρο των λύσεων και αξιολογούνται σύμφωνα με ορισμένο κριτήριο καταλληλότητας (fitness criterion) μετά από κάθε κύκλο χρόνου. Με την πάροδο του χρόνου, τα σωματίδια επιταχύνονται προς τα σωματίδια που έχουν τις καλύτερες τιμές καταλληλότητας (fitness values). Το κύριο πλεονέκτημα μιας τέτοιας προσέγγισης σε σχέση με άλλες στρατηγικές ελαχιστοποίησης, είναι ότι ο μεγάλος αριθμός των μελών που απαρτίζουν το σμήνος σωματιδίων κάνουν την τεχνική εντυπωσιακά ανθεκτική στο πρόβλημα του τοπικού ελάχιστου.

Υποθέστε το ακόλουθο σενάριο: μια ομάδα πουλιών ψάχνει τυχαία τροφή μέσα σε μια περιοχή. Υπάρχει μόνο ένα κομμάτι τροφής στην υπό έρευνα περιοχή. Τα πουλιά δεν ξέρουν πού είναι η τροφή. Αλλά ξέρουν πόσο μακριά είναι η τροφή, καθώς και τις θέσεις των διπλανών τους. Έτσι ποιά είναι η καλύτερη στρατηγική εύρεσης τροφής; Μια αποτελεσματική στρατηγική είναι να ακολουθηθεί το κοντινότερο πουλί στην τροφή.

Ο PSO μαθαίνει από το σενάριο και το χρησιμοποιεί για να λύσει τα προβλήματα βελτιστοποίησης. Στον PSO, κάθε ξεχωριστή λύση είναι σαν ένα «πουλί» στον χώρο αναζήτησης, η οποία καλείται σωματίδιο. Όλα τα σωματίδια έχουν τιμές καταλληλότητας που αξιολογούνται από τη συνάρτηση καταλληλότητας που πρόκειται να βελτιστοποιηθεί, και έχουν ταχύτητες που κατευθύνουν το πέταγμα των σωματιδίων. (Τα σωματίδια κινούνται διαμέσου του χώρου αναζήτησης ακολουθώντας τα σωματίδια με τις καλύτερες λύσεις μέχρι τώρα). Ο PSO αρχικοποιείται με μια ομάδα τυχαίων σωματιδίων (λύσεις) και κατόπιν αναζητά τις βέλτιστες λύσεις ενημερώνοντας την κάθε γενιά.

Κάθε άτομο αντιμετωπίζεται σαν ένα σωματίδιο χωρίς όγκο (ένα σημείο) στο D-διάστατο διάστημα αναζήτησης. Το i-στό σωματίδιο αναπαρίσταται σαν $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$. Σε κάθε γενιά, κάθε σωματίδιο ενημερώνεται από τις ακόλουθες δύο «καλύτερες» τιμές. Η πρώτη είναι η καλύτερη προηγούμενη θέση (η θέση που δίνει την καλύτερη τιμή καταλληλότητας) που ένα σωματίδιο έχει πραγματοποιήσει μέχρι τώρα. Αυτή η τιμή καλείται pBest. Η pBest του i-στού σωματιδίου

αναπαρίσταται σαν $P_i = (p_{i1}, p_{i2}, \dots, p_{iD})$. Σε κάθε επανάληψη, το P διάνυσμα του σωματιδίου με την καλύτερη καταλληλότητα στη γειτονιά, οριζόμενο σαν l ή g , και το P διάνυσμα του τρέχοντος σωματιδίου συνδυάζονται για να ρυθμίσουν την ταχύτητα κατά μήκος κάθε διάστασης, και αυτή η ταχύτητα χρησιμοποιείται στη συνέχεια για να υπολογίσει μια νέα θέση για το σωματίδιο. Το ποσοστό της ρύθμισης της ταχύτητας που επηρεάζεται από την προηγούμενη καλύτερη θέση του ατόμου (P) θεωρείται σαν συστατικό γνώσης, και το ποσοστό που επηρεάζεται από το καλύτερο στη γειτονιά είναι το κοινωνικό συστατικό. Με την προσθήκη του παράγοντα αδράνειας ω , από τους Shi και Eberhart (που έχει εισαχθεί για την εξισορρόπηση της σφαιρικής και τοπικής αναζήτησης), αυτές οι εξισώσεις είναι:

$$v_{id} = \omega * v_{id} + \eta_1 * \text{rand}() * (p_{id} - x_{id}) + \eta_2 * \text{Rand}() * (p_{gd} - x_{id}) \quad (1.1)$$

$$x_{id} = x_{id} + v_{id} \quad (1.2)$$

όπου $\text{rand}()$ και $\text{Rand}()$ είναι δύο τυχαίοι αριθμοί που παράγονται ανεξάρτητα μέσα στο διάστημα $(0,1)$ και η_1 και η_2 είναι δύο παράγοντες εκμάθησης που ελέγχουν την επιρροή των κοινωνικών και γνωστικών συστατικών. Στην εξίσωση (1.1), εάν το άθροισμα στη δεξιά πλευρά υπερβαίνει μια σταθερή τιμή, τότε η ταχύτητα σε εκείνη την διάσταση ορίζεται σαν $\pm V_{max}$.

Κατά συνέπεια, οι ταχύτητες των σωματιδίων κυμαίνονται στο διάστημα $(-V_{max}, V_{max})$ το οποίο λειτουργεί σαν περιορισμός για να ελέγξει τη σφαιρική δυνατότητα εξερεύνησης του σμήνους των σωματιδίων. Κατά συνέπεια, η πιθανότητα των σωματιδίων που αφήνουν το χώρο αναζήτησης μειώνεται. Αυτό δεν γίνεται για να περιορίσει τις τιμές X_i μέσα στο διάστημα $(-V_{max}, V_{max})$, αλλά περιορίζει τη μέγιστη απόσταση που ένα σωματίδιο θα διανύσει κατά τη διάρκεια μιας επανάληψης. Ο βασικός αλγόριθμος PSO όπως περιγράφεται από τον Pomeroy δίνεται παρακάτω:

```

/* set up particles' next location */
for each particle p do {
    for d = 1 to dimensions do {
        p.next(d) = random()
        p.velocity(d) = random(deltaMin, deltaMax)
    }
    p.bestSoFar = initialFitness
}
/* set particles' neighbors */
for each particle p do {
    for n = 1 to numberOfNeighbors do {
        p.neighbor(n) = getNeighbor(p,n)
    }
}
/* run Particle Swarm Optimizer */
while iterations ≤ maxIterations do {
    /* Make the "next locations" current and then*/

```

```

/* test their fitness. */
  for each particle p do {
    for d = 1 to dimensions do {
      p.current(d) = p.next(d)
    }
    fitness = test(p)
    if fitness > p.bestSoFar then do {
      p.bestSoFar = fitness
      for d = 1 to dimensions do {
        p.best(d) = p.current(d)
      }
    }
    if fitness = targetFitness then do {
      ...
      /* e.g., write out solution and quit */
    }
  } /* end of: for each particle p */
  for each particle p do {
    n = getNeighborWithBestFitness(p)
    for d = 1 to dimensions do {
      iFactor = iWeight * random(iMin, iMax)
      sFactor = sWeight * random(sMin, sMax)
      pDelta(d) = p.best(d) - p.current(d)
      nDelta(d) = n.best(d) - p.current(d)
      delta = (iFactor * pDelta(d)) + (sFactor * nDelta(d))
      delta = p.velocity(d) + delta
      p.velocity(d) = constrict(delta)
      p.next(d) = p.current(d) + p.velocity(d)
    }
  } /* end of: for each particle p */
} /* end of: while iterations ≤ maxIterations */
end /* end of main program */
/* Return neighbor n of particle p */
function getNeighbor(p, n) {
  ...
  return neighborParticle
}
/* Return particle in p's neighborhood */
/* with the best fitness */
function getNeighborWithBestFitness(p) {
  ...

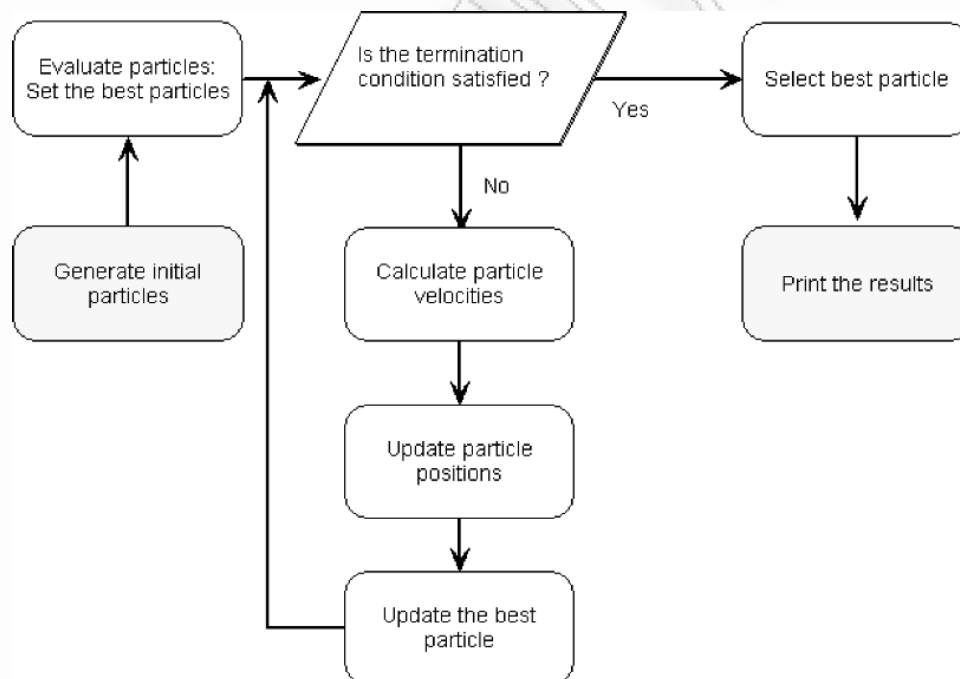
```

```

    return neighborParticle
}
/* Limit the change in a particle's */
/* dimension value */
function constrict(delta) {
    if delta < deltaMin then
        return deltaMin
    else
        if delta > deltaMax then
            return deltaMax
        else
            return delta
}

```

Το βασικό σχήμα του αλγορίθμου PSO παρουσιάζεται στο Σχήμα 5.

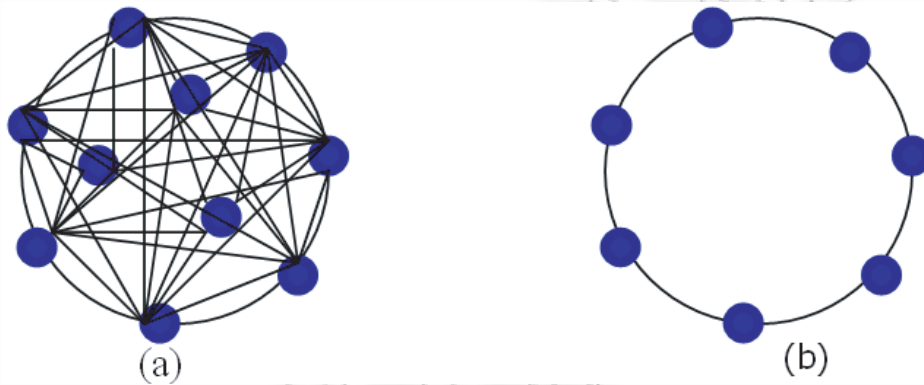


Σχήμα 5: η βασική δομή του αλγορίθμου PSO

Ο PSO αλγόριθμος μπορεί να ειπωθεί ως σύνολο διανυσμάτων των οποίων οι τροχιές ταλαντεύονται γύρω από μια περιοχή καθορισμένη από την προηγούμενη καλύτερη θέση κάθε ατόμου και την καλύτερη θέση κάποιων άλλων ατόμων. Υπάρχουν διαφορετικές τοπολογίες γειτονιάς που χρησιμοποιούνται για να προσδιορίσουν ποιά από τα σωματίδια του σμήνους μπορούν να επηρεάσουν τα υπόλοιπα άτομα. Οι πιο κοινές είναι γνωστές ως *gbest* και *lbest*:

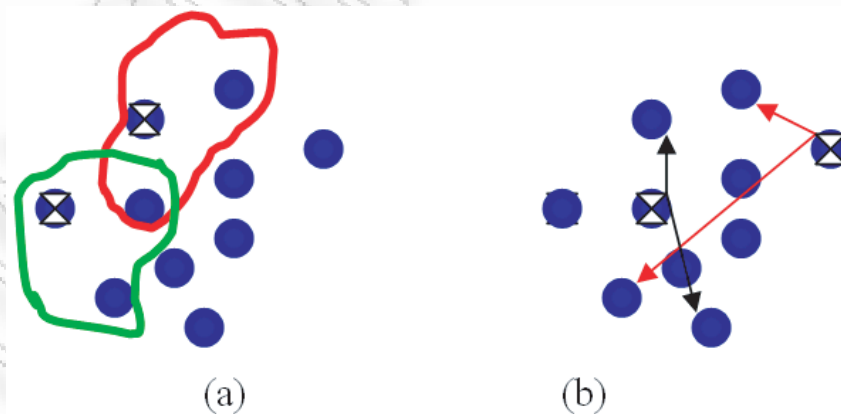
Στο gbest σμήνος, η τροχιά κάθε ατόμου (σωματίδιου) επηρεάζεται από το καλύτερο άτομο ολόκληρου του σμήνους. Υποθέτουμε ότι τα gbest σμήνη συγκλίνουν γρήγορα, καθώς όλα τα σωματίδια προσελκύονται ταυτόχρονα στο καλύτερο μέρος του χώρου αναζήτησης. Εντούτοις, εάν το καθολικό βέλτιστο δεν είναι κοντά στο καλύτερο σωματίδιο, μπορεί να είναι αδύνατο για το σμήνος να εξερευνήσει άλλες περιοχές και, συνεπώς, το σμήνος μπορεί να παγιδευτεί σε ένα τοπικό βέλτιστο.

Στο lbest σμήνος, κάθε άτομο επηρεάζεται από έναν μικρότερο αριθμό γειτόνων του (οι οποίοι φαίνονται σαν παρακαείμενα μέλη της διάταξης του σμήνους). Χαρακτηριστικά, οι lbest γειτονιές περιλαμβάνουν δύο γείτονες: έναν στη δεξιά και έναν στην αριστερή πλευρά (πλέγμα δαχτυλιδιού). Αυτός ο τύπος σμήνους θα συγκλίνει πιο αργά, αλλά μπορεί να εντοπίσει το συνολικό βέλτιστο με μια μεγαλύτερη πιθανότητα. Το lbest σμήνος είναι σε θέση να κινηθεί γύρω από τα τοπικά βέλτιστα, και τα υποσμήνη (subswarms) είναι σε θέση να εξερευνήσουν τα διαφορετικά βέλτιστα. Μια γραφική αναπαράσταση του gbest σμήνους και του lbest σμήνους αντίστοιχα απεικονίζεται στο Σχήμα 6.



Σχήμα 6: Γραφική αναπαράσταση (a) gbest σμήνος, (b) lbest σμήνος

Εάν θεωρήσουμε τις κοινωνικές και γεωγραφικές γειτονιές όπως παρουσιάζονται στο Σχήμα 7, τότε και ο gbest και ο lbest μπορούν να αντιμετωπισθούν ως μορφές κοινωνικών γειτονιών.



Σχήμα 7: Παραδείγματα (a) γεωγραφικής γειτονιάς, (b) κοινωνικής γειτονιάς σε ένα σμήνος

Ο Watts προσδιόρισε δύο παράγοντες που επηρεάζουν την ανταλλαγή πληροφοριών μεταξύ των μελών δικτύων του μικρόκοσμου (π.χ. εξάπλωση επιδημιών):

- ο βαθμός συνδετικότητας (degree of connectivity): η συμπεριφορά κάθε ατόμου θα επηρεαστεί από τη συμπεριφορά των k γειτόνων του.
- το ποσό ομαδοποίησης (amount of clustering): η ομαδοποίηση αναφέρεται στους γείτονες που είναι κοινοί με μερικά άλλα άτομα.

Υπάρχουν αρκετοί τύποι τοπολογιών σμήνους που μπορούν να χρησιμοποιηθούν ανάλογα με το προς επίλυση πρόβλημα. Οι Kennedy και Mendes δοκίμασαν αρκετές τοπολογίες όπως πυραμίδα, αστέρα, von Neumann κ.λ.π. για κάποια προβλήματα βελτιστοποίησης.

1.3 Ant Colonies Optimization

Οι αλγόριθμοι βελτιστοποίησης αποικιών μυρμηγκιών (ACO) εισήχθησαν περίπου το 1990. Αυτοί οι αλγόριθμοι εμπνεύστηκαν από τη συμπεριφορά των αποικιών των μυρμηγκιών. Τα μυρμήγκια είναι κοινωνικά έντομα, που ενδιαφέρονται κυρίως για την επιβίωση της αποικίας παρά για την δική τους επιβίωση. Το ενδιαφέρον είναι η δυνατότητα των μυρμηγκιών να βρουν τη συντομότερη διαδρομή από τη φωλιά τους προς την τροφή. Αυτή η ιδέα ήταν η πηγή των προτεινόμενων αλγορίθμων που εμπνεύστηκαν από τη συμπεριφορά των μυρμηγκιών.

Κατά την αναζήτηση της τροφής, τα μυρμήγκια ερευνούν αρχικά την περιοχή που περιβάλλει τη φωλιά τους με τυχαίο τρόπο. Καθώς κινούνται, τα μυρμήγκια αφήνουν ένα χημικό ίχνος φερομονών στο έδαφος. Τα μυρμήγκια καθοδηγούνται από τη μυρωδιά φερομονών. Τα μυρμήγκια τείνουν να επιλέξουν τις πορείες που χαρακτηρίζονται από την ισχυρότερη συγκέντρωση φερομονών. Όταν ένα μυρμήγκι βρίσκει μια πηγή τροφής, αξιολογεί την ποσότητα και την ποιότητα της τροφής και φέρνει λίγη από αυτή πίσω στη φωλιά. Κατά τη διάρκεια του ταξιδιού της επιστροφής, η ποσότητα φερομόνης που ένα μυρμήγκι αφήνει στο έδαφος μπορεί να εξαρτηθεί από την ποσότητα και την ποιότητα της τροφής. Τα ίχνη φερομονών θα καθοδηγήσουν άλλα μυρμήγκια προς την πηγή της τροφής. Η έμμεση επικοινωνία μεταξύ των μυρμηγκιών μέσω των ιχνών φερομόνης τους επιτρέπει να βρουν τις κοντινότερες διαδρομές μεταξύ της φωλιάς τους και της πηγής της τροφής. Όπως διατυπώθηκε από τους Dorigo et al. τα κύρια βήματα του αλγορίθμου ACO δίνονται παρακάτω:

1. Αρχικοποίηση ίχνους φερομονών
2. Κατασκευή λύσης που χρησιμοποιεί το ίχνος των φερομονών. Κάθε μυρμήγκι κατασκευάζει μια πλήρη λύση στο πρόβλημα σύμφωνα με έναν πιθανολογικό κανόνα μετάβασης κατάστασης
3. Ο κανόνας μετάβασης κατάστασης εξαρτάται κυρίως από την κατάσταση της φερομόνης
4. Ενημέρωση ίχνους φερομονών.

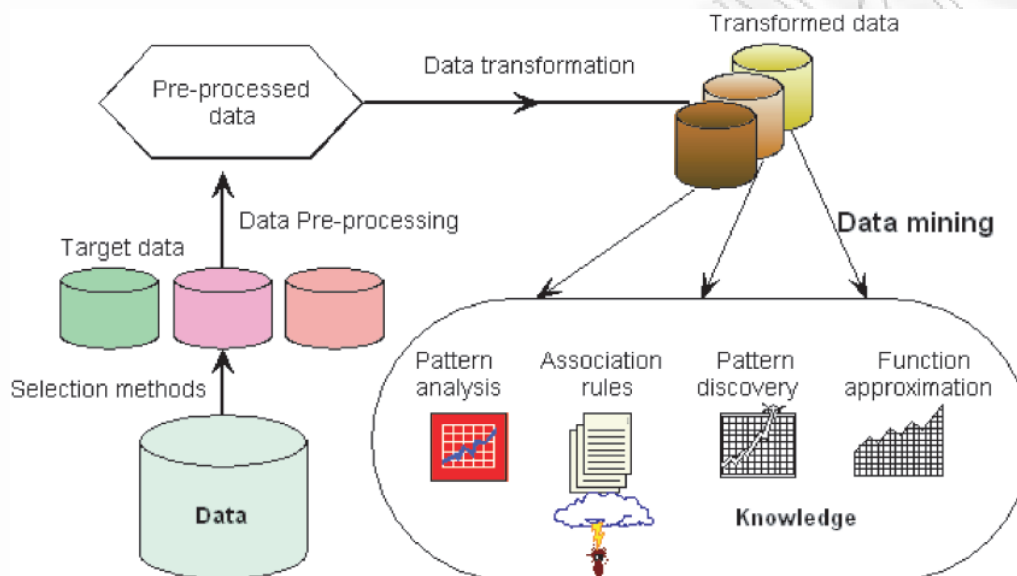
Ένας γενικός κανόνας ενημέρωσης φερομονών εφαρμόζεται σε δύο φάσεις. Κατ' αρχάς, μια φάση εξάτμισης κατά την οποία ένα μέρος της φερομόνης εξατμίζεται και έπειτα μια φάση ενίσχυσης κατά την οποία κάθε μυρμήγκι καταθέτει ένα ποσό φερομόνης που είναι ανάλογο προς την καταλληλότητα της λύσης του. Αυτή η διαδικασία επαναλαμβάνεται έως ότου επιτευχθεί μια συνθήκη τερματισμού.

1.3.1 Εξόρυξη γνώσης από δεδομένα

Ιστορικά, στην έννοια εύρεσης χρήσιμων προτύπων στα δεδομένα έχει δοθεί μια ποικιλία ονομάτων συμπεριλαμβανομένων της εξόρυξης γνώσης από δεδομένα, της εξαγωγής γνώσης, της ανάκτησης πληροφοριών, και της επεξεργασίας προτύπων δεδομένων. Η εξόρυξη γνώσης από δεδομένα είναι η εφαρμογή των συγκεκριμένων αλγορίθμων για την εξαγωγή προτύπων από δεδομένα. Τα πρόσθετα

βήματα στη διαδικασία KDD (Knowledge Discovery and Data Mining), όπως η επιλογή των δεδομένων, ο καθαρισμός των δεδομένων, η ενσωμάτωση της κατάλληλης προγενέστερης γνώσης, και η κατάλληλη ερμηνεία των αποτελεσμάτων είναι σημαντικά για να εξασφαλίσουν ότι παράγεται χρήσιμη γνώση από τα δεδομένα.

Παρακάτω θα περιγραφούν μερικά από τα βασικά βήματα της διαδικασίας εξόρυξης δεδομένων όπως φαίνονται στο Σχήμα 8.



Σχήμα 8: Βήματα της διαδικασίας ανακάλυψης της γνώσης

1. Ανάπτυξη και κατανόηση του πεδίου εφαρμογής, της σχετικής προγενέστερης γνώσης, και προσδιορισμός του στόχου της διαδικασίας KDD.
2. Δημιουργία του στόχου συνόλου δεδομένων
3. Καθαρισμός και προεπεξεργασία των δεδομένων: βασικές διαδικασίες όπως η αφαίρεση του θορύβου και η διαχείριση ελλειπών πεδίων δεδομένων.
4. Μείωση και προβολή δεδομένων: εύρεση των χρήσιμων χαρακτηριστικών γνωρισμάτων για την παρουσίαση των δεδομένων σε σχέση με τον σκοπό. Χρήση της μείωσης των διαστάσεων ή μεθόδων μετασχηματισμού για να μειωθεί ο ενεργός αριθμός των υπό εξέταση μεταβλητών ή για να βρεθεί μια αμετάβλητη παρουσίαση των δεδομένων.
5. Ταίριασμα των στόχων της διαδικασίας KDD με μια συγκεκριμένη μέθοδο εξόρυξης δεδομένων: Αν και τα όρια μεταξύ της πρόβλεψης και της περιγραφής δεν είναι σαφή, η διάκριση είναι χρήσιμη για την κατανόηση του γενικού στόχου ανακάλυψης.

Οι στόχοι της ανακάλυψης γνώσης επιτυγχάνονται μέσω των ακόλουθων μεθόδων εξόρυξης δεδομένων:

- Ομαδοποίηση (Clustering): προσδιορισμός ενός πεπερασμένου συνόλου κατηγοριών ή συστάδων που περιγράφουν τα δεδομένα.
- Άθροισμα (Summation): εύρεση μιας συμπαγούς περιγραφής για το υποσύνολο των δεδομένων, π.χ. η δημιουργία περίληψης για τους κανόνες συσχέτισης και η χρήση τεχνικών απεικόνισης πολλών μεταβλητών.
- Μοντελοποίηση εξάρτησης (Dependency modeling): εύρεση ενός προτύπου που περιγράφει σημαντικές εξαρτήσεις μεταξύ των μεταβλητών.

- Παλινδρόμηση (Regression): εκμάθηση μιας συνάρτησης που αντιστοιχεί ένα δεδομένο σε μια μεταβλητή πρόβλεψης πραγματικής τιμής και η ανακάλυψη των συναρτησιακών σχέσεων μεταξύ των μεταβλητών.
- Ταξινόμηση (Classification): εκμάθηση μιας συνάρτησης που ταξινομεί ένα στοιχείο δεδομένων σε μια από τις προκαθορισμένες κατηγορίες.
- Ανίχνευση αλλαγής και απόκλισης (Change and Deviation Detection): ανακάλυψη των σημαντικότερων αλλαγών στα δεδομένα από τις προηγούμενα μετρημένες ή κανονικοποιημένες τιμές.

1.4 Εφαρμογές της Νοημοσύνης Σμήνους στην εξόρυξη γνώσης από δεδομένα

1.4.1 Particle Swarm Optimization και εξόρυξη γνώσης από δεδομένα

Η εξόρυξη γνώσης από δεδομένα και η βελτιστοποίηση σμήνους σωματιδίων μπορεί να φαίνεται ότι δεν έχουν πολλές κοινές ιδιότητες. Εντούτοις, μπορούν να χρησιμοποιηθούν μαζί για να διαμορφώσουν μια μέθοδο που συχνά οδηγεί σε αποτέλεσμα, ακόμα και όταν άλλες μέθοδοι θα ήταν πάρα πολύ ακριβές ή δύσκολες να εφαρμοστούν. Ο Omran (2005), (2002) έχει χρησιμοποιήσει τις μεθόδους βελτιστοποίησης σμήνους σωματιδίων για αναγνώριση προτύπων και επεξεργασία εικόνας. Μια νέα μέθοδος ομαδοποίησης βασισμένη στον PSO προτείνεται και εφαρμόζεται σε ταξινόμηση χωρίς επιτήρηση και στην κατάτμηση εικόνας.

Η εξόρυξη οπτικής γνώσης από δεδομένα μέσω της κατασκευής χώρων εικονικής πραγματικότητας για την παρουσίαση των δεδομένων και της γνώσης, περιλαμβάνει την επίλυση προβλημάτων βελτιστοποίησης. Ο Valdes (2004) εισήγαγε μια υβριδική τεχνική βασισμένη στη βελτιστοποίηση σμήνους σωματιδίων (PSO) σε συνδυασμό με κλασσικές μεθόδους βελτιστοποίησης. Αυτή η προσέγγιση εφαρμόζεται σε δεδομένα πολλών διαστάσεων προκειμένου να γίνει κατανοητή η δομή των ακατέργαστων και των επεξεργασμένων δεδομένων. Πειράματα με σύνολα δεδομένων που αντιστοιχούν στην ασθένεια του Alzheimer έδειξαν ότι υψηλής ποιότητας οπτική παρουσίαση μπορεί να ληφθεί με το συνδυασμό του PSO με τις κλασσικές μεθόδους βελτιστοποίησης. Η συμπεριφορά μερικών από τις παραμέτρους που ελέγχουν την εξέλιξη των σμηνών μελετήθηκε επίσης.

Ο Sousa et al. (2003), (2004) έχουν προτείνει τη χρήση του PSO ως εργαλείο για εξόρυξη δεδομένων. Προκειμένου να αξιολογηθεί η χρησιμότητα του PSO για την εξόρυξη γνώσης από δεδομένα, χρησιμοποιήθηκε μια εμπειρική σύγκριση της απόδοσης τριών παραλλαγών PSO με έναν άλλο εξελικτικό αλγόριθμο (γενετικός αλγόριθμος), για την ανακάλυψη κανόνων με σκοπό την ταξινόμηση. Τέτοιοι στόχοι θεωρούνται βασικά εργαλεία για τα συστήματα υποστήριξης απόφασης σε μια ευρεία περιοχή, που κυμαίνεται από τη βιομηχανία και το εμπόριο, έως στρατιωτικούς και επιστημονικούς τομείς. Οι πηγές των δεδομένων που χρησιμοποιούνται για την πειραματική δοκιμή θεωρούνται ως de facto πρότυπα για την ταξινόμηση βάσει της αξιοπιστίας των αλγορίθμων ανακάλυψης κανόνων. Τα αποτελέσματα που αποκτήθηκαν σε αυτές τις περιοχές έδειξαν ότι οι αλγόριθμοι PSO είναι ανταγωνιστικοί με άλλες εξελικτικές τεχνικές, και μπορούν να εφαρμοστούν επιτυχώς σε περισσότερο απαιτητικά πεδία προβλημάτων.

Τα Συστήματα Σύστασης είναι νέοι τύποι εργαλείων λογισμικού βασισμένοι στο Διαδίκτυο, σχεδιασμένα για να βοηθήσουν τους χρήστες να βρουν τον δρόμο τους μέσω των σημερινών σύνθετων on-line καταστημάτων και ιστοχώρων ψυχαγωγίας. Οι Ujjin και Bentley (2002), (2003) έχουν περιγράψει έναν νέο σύστημα σύστασης, το οποίο υιοθετεί έναν αλγόριθμο βελτιστοποίησης σμήνους σωματιδίων (PSO) για να μάθει τις προσωπικές προτιμήσεις των χρηστών και να παράσχει προσαρμοσμένες συστάσεις. Τα πειράματα πραγματοποιούνται για να παρατηρήσουν την απόδοση του συστήματος και τα αποτελέσματα συγκρίνονται έναντι εκείνων που λαμβάνονται από ένα σύστημα σύστασης γενετικού

αλγορίθμου (GA) και ένα τυποποιημένο, μη-προσαρμοστικό σύστημα βασισμένο στον αλγόριθμο Pearson (Breese et al, 1998).

Μια άλλη πολύ σημαντική εφαρμογή PSO είναι στο πεδίο των cascading classifiers. Οι cascading ταξινομητές έχουν χρησιμοποιηθεί τα τελευταία χρόνια για την επίλυση προβλημάτων αναγνώρισης προτύπων. Τα κύρια κίνητρα πίσω από μια τέτοια στρατηγική είναι η βελτίωση της ακρίβειας ταξινόμησης και η μείωση της πολυπλοκότητας.

Οι Settles και Rylander (2002) έχουν προτείνει μια μέθοδο PSO για την εκπαίδευση νευρωνικού δικτύου. Οι Chen και Abraham (2006) ερεύνησαν πώς η φαινομενικά χαοτική συμπεριφορά των χρηματιστηρίων θα μπορούσε να αντιπροσωπευθεί καλά χρησιμοποιώντας διάφορες soft τεχνικές υπολογισμού. Οι συντάκτες εξέτασαν τον flexible neural tree αλγόριθμο, ένα wavelet νευρωνικό δίκτυο, ένα τοπικό, γραμμικό wavelet νευρωνικό δίκτυο και τελικά ένα feed-forward τεχνητό νευρωνικό δίκτυο. Οι παράμετροι των διαφορετικών τεχνικών εκμάθησης βελτιστοποιήθηκαν από την PSO προσέγγιση. Τα αποτελέσματα του πειράματος αποκαλύπτουν ότι ο PSO θα μπορούσε να διαδραματίσει έναν σημαντικό ρόλο για να τελειοποιήσει τις παραμέτρους για τη βέλτιστη απόδοση.

Ο καρκίνος του μαστού είναι μια από τις σημαντικότερες αιτίες θανάτου στις γυναίκες. Διάφορες τεχνικές τεχνητής νοημοσύνης έχουν χρησιμοποιηθεί για να βελτιώσουν τις διαδικασίες διαγνώσεων και για να βοηθήσουν τις προσπάθειες των γιατρών. Οι Chen και Abraham (2005) παρουσίασαν μια μελέτη για την ανίχνευση του καρκίνου του μαστού που χρησιμοποιεί ένα Flexible Neural Tree (FNT), ένα Neural Network (NN), ένα Wavelet Neural Network (WNN) και τον συνδυασμό τους. Για το μοντέλο FNT, ένας βασισμένος σε δομή δέντρου εξελικτικός αλγόριθμος και ένα ένας PSO χρησιμοποιούνται για να βρεθεί ένα βέλτιστο FNT. Για το NN και το WNN, ο PSO υιοθετείται για να βελτιστοποιήσει τις ελεύθερες παραμέτρους. Η απόδοση κάθε προσέγγισης αξιολογείται χρησιμοποιώντας το σύνολο δεδομένων καρκίνου του μαστού. Τα αποτελέσματα προσομοίωσης δείχνουν ότι το αποκτηθέν πρότυπο FNT έχει μικρότερο αριθμό μεταβλητών με μειωμένο αριθμό χαρακτηριστικών γνωρισμάτων εισόδου και χωρίς σημαντική μείωση της ακρίβειας ανίχνευσης. Η γενική ακρίβεια θα μπορούσε να βελτιωθεί με τη χρησιμοποίηση μιας συνολικής προσέγγισης με μια μέθοδο ψηφοφορίας.

Οι Chen et al. (2006) πρότειναν μια εξελικτική διαδικασία για τη σχεδίαση ιεραρχικού ή πολυεπίπεδου fuzzy συστήματος (Takagi-Sugeno Fuzzy Systems - TS-FS).

Η ανάλυση ομάδων (Cluster analysis) έχει γίνει μια σημαντική τεχνική στη διερευνητική ανάλυση δεδομένων, στην αναγνώριση προτύπων, στη μηχανική μάθηση, στους νευρωνικούς υπολογισμούς, και σε άλλα πεδία εφαρμοσμένης μηχανικής. Η ομαδοποίηση στοχεύει στον προσδιορισμό και στην εξαγωγή σημαντικών ομάδων στα δεδομένα. Οι τέσσερις κύριες κατηγορίες αλγορίθμων ομαδοποίησης είναι οι μέθοδοι διαμερισμού, οι ιεραρχικές μέθοδοι, η βασισμένη στην πυκνότητα ομαδοποίηση και η βασισμένη στο πλέγμα ομαδοποίηση. Η ομαδοποίηση εγγράφων είναι μια θεμελιώδης λειτουργία που χρησιμοποιείται στη μη επιτηρούμενη οργάνωση εγγράφων, στην αυτόματη εξαγωγή θέματος, και στην ανάκτηση πληροφοριών. Γρήγοροι και υψηλής ποιότητας αλγόριθμοι ομαδοποίησης εγγράφων διαδραματίζουν έναν σημαντικό ρόλο στην αποτελεσματική πλοήγηση, την περιήληψη, και την οργάνωση των πληροφοριών. Στον τομέα της ομαδοποίησης, ο αλγόριθμος K-means είναι ο γενικότερα χρησιμοποιούμενος αλγόριθμος για να βρεθεί ένα τμήμα που ελαχιστοποιεί το μέτρο του μέσου τετραγωνικού λάθους (MSE). Αν και ο K-means είναι ένας εκτενώς χρήσιμος αλγόριθμος ομαδοποίησης, υποφέρει από διάφορα μειονεκτήματα. Η αντικειμενική συνάρτηση του K-means δεν είναι κυρτή και ως εκ τούτου μπορεί να περιέχει τοπικά ελάχιστα. Συνεπώς, ελαχιστοποιώντας την αντικειμενική συνάρτηση, υπάρχει η πιθανότητα κολλήματος στα τοπικά ελάχιστα (επίσης στα τοπικά μέγιστα και στο σημείο καμπής) (Selim SZ, Ismail MA, 1984). Η απόδοση του αλγόριθμου K-means εξαρτάται από την αρχική επιλογή των κέντρων των ομάδων. Εκτός αυτού, ο ευκλείδειος κανόνας είναι ευαίσθητος στο θόρυβο ή στα outliers. Ως εκ τούτου ο αλγόριθμος K-means πρέπει να επηρεάζεται από το θόρυβο και τα outliers (Wu KL, Yang MS (2002)), (Jones G, Robertson A, Santimetvirul C, Willett P 1995). Εκτός από τον αλγόριθμο K-means, διάφοροι αλγόριθμοι, όπως ο γενετικός αλγόριθμος (GA) (Jones G, Robertson

A, Santimetvirul C, Willett P 1995), (Raghavan VV, Birchard K 1979) και οι Self-Organizing χάρτες (SOM) (Merkel D 2002), έχουν χρησιμοποιηθεί για την ομαδοποίηση εγγράφων.

Πράκτορες σε σμήνη σε δίκτυα φυσικά καταναμημένων κόμβων επεξεργασίας μπορούν να χρησιμοποιηθούν για την απόκτηση δεδομένων και για εφαρμογές ελέγχου. Χρησιμοποιείται μια αρχιτεκτονική για ενεργά συστήματα επιτήρησης στα οποία οι απλοί κινητοί πράκτορες επεξεργάζονται συλλογικά τα δεδομένα πραγματικού χρόνου από τις ετερογενείς πηγές στην προέλευση ή κοντά στην προέλευση των δεδομένων.

Οι Ye και Chen (2005) εισήγαγαν μια εξελικτική PSO βασισμένη στη μάθηση μέθοδο για τη βέλτιστη συσταδοποίηση N σημείων δεδομένων σε K συστάδες. Ο υβριδικός PSO και K -means αλγόριθμος καλείται Alternative KPSO-clustering (AKPSO). Αυτός αναπτύχθηκε για την αυτόματη ανίχνευση των κέντρων των συστάδων στα σύνολα δεδομένων γεωμετρικής δομής. Ο αλγόριθμος AKPSO θεωρείται ότι βελτιώνει τον παραδοσιακό K -means αλγόριθμο για να αντιμετωπίσει σύνολα δεδομένων διάφορων δομών. Αποτελέσματα προσομοίωσης σε σύγκριση με μερικές γνωστές μεθόδους ομαδοποίησης κατέδειξαν την ευρωστία και την αποδοτικότητα της νέας μεθόδου AKPSO.

1.4.2 Ant Colony Optimization και εξόρυξη γνώσης από δεδομένα

Υπάρχουν πολλοί αλγόριθμοι εμπνευσμένοι από τις αποικίες των μυρμηγκιών. Τα μυρμηγκία συλλέγουν αντικείμενα για να διαμορφώσουν σωρούς (συγκέντρωση νεκρών σωμάτων). Ο βασικός μηχανισμός πίσω από αυτόν τον τύπο φαινομένου συλλογής είναι ότι οι μικρές συστάδες των αντικειμένων αυξάνονται με την προσέλευση των μυρμηγκιών να αποθέσουν περισσότερα αντικείμενα. Είναι αυτή η θετική ανατροφοδότηση που οδηγεί στο σχηματισμό ολοένα μεγαλύτερων συστάδων.

Η γενική ιδέα για τη συσταδοποίηση των δεδομένων είναι ότι τα απομονωμένα δεδομένα πρέπει να παρθούν και να αποτεθούν σε κάποια άλλη θέση, όπου περισσότερα δεδομένα του ίδιου τύπου είναι παρόντα. Οι Ramos et al. (2002) πρότειναν τον A-CLUSTER αλγόριθμο για να προσομοιώσουν όσο το δυνατόν τις συμπεριφορές που μοιάζουν με αυτές των πραγματικών μυρμηγκιών. Υπό αυτή την έννοια, οι βιο-εμπνευσμένες χωρικές πιθανότητες μετάβασης ενσωματώνονται στο σύστημα, αποφεύγοντας τους τυχαία κινούμενους πράκτορες, οι οποίοι ενθαρρύνουν τον καταναμημένο αλγόριθμο να εξερευνησει περιοχές προφανώς χωρίς ενδιαφέρον. Η στρατηγική επιτρέπει στα μυρμηγκία να βρουν τις συστάδες των αντικειμένων με έναν προσαρμοστικό τρόπο.

Προκειμένου να μοντελοποιηθεί η συμπεριφορά των μυρμηγκιών που συνδέεται με τη λήψη και την ρίψη αντικειμένων προτάθηκε η χρήση συνδυασμών διαφορετικών καταωφλίων απόκρισης. Υπάρχουν δύο σημαντικοί παράγοντες που πρέπει να επηρεάσουν οποιαδήποτε τοπική δράση που αναλαμβάνεται από τον ant-like πράκτορα: ο αριθμός των αντικειμένων στη γειτονιά του, και η ομοιότητά τους. Οι Lumer και Faieta (1994) χρησιμοποίησαν μια μέση ομοιότητα, αναμιγνύοντας τις αποστάσεις μεταξύ των αντικειμένων με τον αριθμό τους, ενσωματώνοντας τον ταυτόχρονα σε μια συνάρτηση καταωφλίου απόκρισης, όπως ο αλγόριθμος που προτείνεται από τους Deneubourg et al. (1991).

Οι Admane et al. (2006), παρουσίασαν τον AntPart, ο οποίος είναι μια αποκλειστική μη επιτηρούμενη τεχνική ταξινόμησης που εμπνέεται από τη συμπεριφορά ενός ιδιαίτερου είδους μυρμηγκιών αποκαλούμενου Apicalis Pachycondyla. Οι αποδόσεις αυτής της μεθόδου συγκρίθηκαν με εκείνες τριών άλλων, που εμπνέονται επίσης από την κοινωνική συμπεριφορά των μυρμηγκιών: AntClass, AntTree και AntClust.

Οι Kuo et al. (2005) πρότειναν τη μέθοδο συσταδοποίησης ant K -means (AK). Ο AK αλγόριθμος τροποποιεί τον K -means καθώς τοποθετεί τα αντικείμενα σε μια συστάδα με την πιθανότητα, η οποία ενημερώνεται από τη φερομόνη, ενώ ο κανόνας της ενημέρωσης της φερομόνης είναι σύμφωνος με το σύνολο μέσα στη μεταβλητότητα των συστάδων (TWCV).

Οι Weng et al. (2006) πρότειναν έναν αλγόριθμο κατάτμησης χρονοσειρών βασισμένο στον αλγόριθμο βελτιστοποίησης αποικιών μυρμηγκιών (ACO) για να εκθέσουν την μεταβλητότητα των δεδομένων των χρονοσειρών. Οι ερευνητές χρησιμοποίησαν την Bottom-Up μέθοδο, η οποία δίνει καλά

αποτελέσματα για την κατάτμηση χρονοσειράς. Το ερευνητικό αποτέλεσμα δείχνει ότι η κατάτμηση χρονοσειράς εκτελεσμένη από τον αλγόριθμο ACO όχι μόνο αυτόματα προσδιορίζει τον αριθμό των τμημάτων, αλλά το κόστος κατάτμησής του ήταν χαμηλότερο από αυτό της κατάτμησης χρονικής σειράς με την Bottom-Up μέθοδο.

Οι Shelokar et al.(2004) ανέπτυξαν μια metaheuristic βελτιστοποίηση αποικιών μυρμηγκιών σαν μια μέθοδο μηχανικής μάθησης βασισμένη σε κανόνες, που λέγεται σύστημα ταξινομητή αποικίας μυρμηγκιών.

Η εξόρυξη γνώσης από τη χρήση του Παγκόσμιου Ιστού προσπαθεί να ανακαλύψει τη χρήσιμη γνώση από τα δευτερεύοντα δεδομένα που λαμβάνονται από τις αλληλεπιδράσεις των χρηστών με το Web. Η εξόρυξη γνώσης από τη χρήση του Web έχει γίνει πολύ κρίσιμη για την αποτελεσματική διαχείριση των τόπων του Web, για τη δημιουργία προσαρμόσιμων τόπων Web, για επιχειρηματικές υπηρεσίες, καθώς και υπηρεσίες υποστήριξης, για εξατομίκευση, για ανάλυση της κίνησης των δικτύων κ.λ.π.. Οι Abraham και Ramos (2003) πρότειναν έναν αλγόριθμο συσταδοποίησης μυρμηγκιών για την αποκάλυψη των προτύπων χρήσης του Ιστού (συστάδες δεδομένων) και μια γραμμική γενετική προγραμματιστική προσέγγιση για την ανάλυση των τάσεων των επισκεπτών. Τα εμπειρικά αποτελέσματα σαφώς δείχνουν ότι η συσταδοποίηση αποικίας μυρμηγκιών δουλεύει καλά όταν συγκρίνεται με έναν self organizing χάρτη (για τη συσταδοποίηση προτύπων χρήσης του Ιστού).

ΚΕΦΑΛΑΙΟ 2

2. Συστήματα Σύστασης

Η νέα χιλιετία είναι η εποχή της αφθονίας των πληροφοριών. Τη δεκαετία του 1990 είδαμε μια έκρηξη των τεχνολογιών πληροφορίας και ψυχαγωγίας, και, ως εκ τούτου, των επιλογών που αντιμετωπίζει ένα άτομο. Οι άνθρωποι μπορεί να επιλέγουν ανάμεσα από εκατοντάδες κανάλια τηλεόρασης, χιλιάδες βίντεο, εκατομμύρια βιβλία, CD και πολυμέσα, αλληλεπιδραστικά έγγραφα στο World Wide Web και φαινομενικά αμέτρητα άλλα καταναλωτικά αγαθά που παρουσιάζονται σε καταλόγους ή διαφημίσεις. Ο Παγκόσμιος Ιστός προσφέρει μυριάδες δυνατότητες – εκτός από αλληλεπιδραστικά έγγραφα, υπάρχουν άπειρες συνομιλίες για να συμμετάσχει κανείς, καθώς και αγαθά για να αγοράσει. Όχι μόνο υπάρχει ένας μεγάλος αριθμός δυνατοτήτων, αλλά διαφέρουν πολύ και στην ποιότητα. Η αξιολόγηση όλων αυτών των εναλλακτικών λύσεων, ωστόσο, εξακολουθεί να απαιτεί σχετικά τον ίδιο χρόνο και την ίδια προσπάθεια όπως πάντα. Η προσοχή μας παραμένει όπως ήταν – η έκρηξη της πληροφορίας δεν έχει συνοδευτεί από ένα άλμα στην ανθρώπινη εξέλιξη. Ως εκ τούτου, τα άτομα δεν μπορούν να ελπίζουν ότι θα αξιολογήσουν όλες τις διαθέσιμες επιλογές μόνοι τους, εκτός εάν το θέμα του ενδιαφέροντός τους είναι αυστηρά περιορισμένο.

Συνεπώς τί μπορούμε να κάνουμε; Όταν οι άνθρωποι έχουν να κάνουν μια επιλογή χωρίς προσωπική γνώση των εναλλακτικών λύσεων, η φυσική πορεία της δράσης είναι να στηρίζονται στην εμπειρία και στις γνώμες των άλλων. Επιδιώκουμε συστάσεις από άτομα που είναι εξοικειωμένα με τις επιλογές που αντιμετωπίζουμε, που έχουν υπάρξει χρήσιμα κατά το παρελθόν, τις απόψεις των οποίων εκτιμάμε, ή που είναι αναγνωρισμένα σαν εμπειρογνώμονες. Μπορεί να απευθυνόμαστε σε φίλους ή συναδέλφους, στον ιδιοκτήτη του γειτονικού βιβλιοπωλείου, στις κριτικές ταινιών των εφημερίδων ή των περιοδικών, ή στις αξιολογήσεις προϊόντων των ενώσεων καταναλωτών. Και μπορεί να βρίσκουμε την κοινωνική διαδικασία της συνάντησης και συνομιλίας με άτομα που μοιράζονται τα ενδιαφέροντά μας εξίσου σημαντική με τις συστάσεις που λαμβάνουμε.

Σήμερα ολοένα και περισσότεροι άνθρωποι στρέφονται προς τα υπολογιστικά συστήματα σύστασης, τα οποία προέκυψαν σαν απάντηση στις τεχνολογικές δυνατότητες και τις ανθρώπινες ανάγκες που δημιουργούνται από το World Wide Web. Αυτά τα συστήματα αποσκοπούν στο να μεσολαβήσουν, να υποστηρίξουν ή να αυτοματοποιήσουν την καθημερινή διαδικασία ανταλλαγής συστάσεων.

Εδώ θα παρουσιάσουμε ένα εννοιολογικό πλαίσιο για προβληματισμό σχετικά με τα συστήματα σύστασης που βασίζεται σε καθημερινά παραδείγματα σύστασης και προσδιορίζει βασικές έννοιες και θέματα σχεδιασμού. Το μεγαλύτερο μέρος του παρόντος κεφαλαίου είναι αφιερωμένο στην επισκόπηση των σχετικά πολλών διαφορετικών προσεγγίσεων για τα συστήματα σύστασης και στην ανάλυσή τους. Τέλος, προτείνουμε μια σειρά από προκλήσεις και ευκαιρίες για την έρευνα και τις εφαρμογές. Δύο κύριες προκλήσεις είναι:

- (1) να βοηθήσουμε τα άτομα να σχηματίσουν κοινότητες ενδιαφέροντος σεβόμενοι την ιδιωτικότητά τους και
- (2) να αναπτύξουμε αλγορίθμους που να συνδυάζουν πολλούς τύπους πληροφορίας για τον υπολογισμό συστάσεων.

2.1 Σύσταση: Αρχές και παραδείγματα

Όλοι μπορούμε να σκεφτούμε παραδείγματα σύστασης, όπως η ανάγνωση κριτικής ταινιών σε ένα περιοδικό για να αποφασίσουμε ποιά ταινία θα δούμε. Ή μπορούμε να θυμηθούμε επισκέψεις στο βιβλιοπωλείο της περιοχής μας, όπου συζητήσαμε με τον βιβλιοπώλη για τα ενδιαφέροντά μας και την τρέχουσα διάθεσή μας, και στη συνέχεια μας πρότεινε κάποια βιβλία που πιθανώς θα μας άρεσαν. Τέλος, μπορούμε να θυμηθούμε να περπατάμε στην περιοχή μας και να παρατηρήσουμε ότι ένα συγκεκριμένο καφέ είναι πάντα γεμάτο κόσμο. Πιστεύουμε ότι η δημοτικότητα του είναι καλό σημάδι, έτσι ώστε να αποφασίσουμε να το επισκεφθούμε.

Η σκέψη πάνω σ' αυτά τα παραδείγματα βοηθά να διευκρινιστεί η έννοια της σύστασης. Ένα άτομο έρχεται αντιμέτωπο με μια απόφαση, η οποία για τους σκοπούς μας είναι μια επιλογή ανάμεσα από ένα σύμπαν εναλλακτικών λύσεων. Το σύμπαν είναι συνήθως αρκετά μεγάλο και κατά πάσα πιθανότητα το άτομο δεν γνωρίζει ποιές είναι όλες οι εναλλακτικές λύσεις, χώρια το πώς μπορεί να επιλέξει μεταξύ τους. Εάν το πρόσωπο δεν έχει επαρκή προσωπική γνώση για να κάνει την επιλογή, μπορεί να ζητήσει συστάσεις από άλλους. Η σύσταση, ως εκ τούτου, είναι μια επικοινωνιακή πράξη.

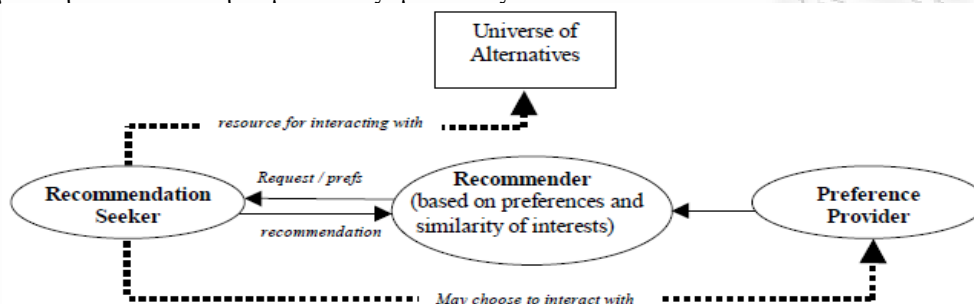
Η σύσταση βασίζεται στις προτιμήσεις αυτού που την δίνει (και ίσως και αυτού που την ζητάει, καθώς και άλλων ατόμων). Για μας, μια προτίμηση είναι μια μεμονωμένη ψυχική κατάσταση σχετικά με ένα υποσύνολο στοιχείων από το σύμπαν των εναλλακτικών λύσεων. Τα άτομα διαμορφώνουν προτιμήσεις, με βάση την εμπειρία τους με τα σχετικά είδη, όπως η ακρόαση μουσικής, η παρακολούθηση ταινιών, η δοκιμή γεύσεων κ.λπ. Για παράδειγμα, κάποιος μπορεί να προτιμά παγωτό βανίλια ή φράουλα (μεταξύ γεύσεων παγωτού), Bach, Mozart και Haydn (μεταξύ κλασικών συνθετών), Joel Coen και Francis Ford Coppola (μεταξύ σύγχρονων σκηνοθετών ταινιών). Φυσικά, οι προτιμήσεις μπορεί να είναι πιο περίπλοκες: ίσως κάποιος να προτιμά κάτι σε σχέση με κάτι άλλο (The Simpsons περισσότερο από τα X-Files) ή ακόμη να σκέφτεται με όρους ενός συστήματος βαθμολόγησης (π.χ. σε μια κλίμακα από 1 έως 10, το Desire του Bob Dylan αξιολογείται με 9).

Μια σύσταση μπορεί να απευθύνεται σε συγκεκριμένα άτομα ή να είναι διαθέσιμη σε όποιον ενδιαφέρεται. Για το άτομο που τη λαμβάνει, μια σύσταση είναι ένας πόρος που το βοηθά να κάνει μια επιλογή από το σύμπαν των εναλλακτικών λύσεων. Η σύσταση λειτουργεί σαν μια προβολή ή σαν φίλτρο επάνω στο συχνά μη προσβάσιμο σύμπαν. Η σύσταση μπορεί να βασίζεται όχι μόνο στις προτιμήσεις αυτού που την κάνει, αλλά και στις προτιμήσεις αυτού που την ζητάει. Για παράδειγμα, εάν πρόκειται να συστήσουμε βιβλία σε κάποιον, μπορούμε να βρούμε ποιά είδη του αρέσουν (π.χ. επιστημονικής φαντασίας) και ακόμη ποιά βιβλία έχει πραγματικά απολαύσει. Στη συνέχεια μπορούμε να προτείνουμε βιβλία που να είναι καλά (κατά τη γνώμη μας) και θα ανταποκρίνονται στις προτιμήσεις του. Μπορούμε ακόμη να συστήσουμε βιβλία με βάση τις προτιμήσεις των άλλων: ίσως δεν είμαστε οπαδοί της επιστημονικής φαντασίας, αλλά έχουμε φίλους που είναι, έτσι, μπορούμε να προτείνουμε σε συστάσεις που βασίζονται σε ό,τι αρέσει στους φίλους μας. Περαιτέρω, μπορούμε να φέρουμε τον ερωτώμενο σε επαφή με άλλα άτομα που έχουν τα ίδια ενδιαφέροντα. Τέλος, μια σύσταση μπορεί να περιλαμβάνει επεξηγηματικό υλικό που να βοηθά αυτόν που ζητά τη σύσταση να την αξιολογήσει.

2.1.1 Ένα μοντέλο της διαδικασίας σύστασης

Το Σχήμα 1 (Terveen και Hill 2001) συνοψίζει αυτές τις έννοιες και τις τοποθετεί σε ένα γενικό μοντέλο σύστασης. Κάποιος δύναται να ζητήσει μια σύσταση, ή ένα σύστημα σύστασης μπορεί να οδηγήσει σε συστάσεις χωρίς να έχει δεχτεί κάποια ερώτηση. Οι ζητούντες συστάσεις μπορούν να συνεισφέρουν τις δικές τους προτιμήσεις, ή τα συστήματα συστάσεων μπορεί να ρωτήσουν γι' αυτές. Με βάση ένα σύνολο γνωστών προτιμήσεων – τη δική του, αυτού που ζητάει τη σύσταση, άλλων ατόμων, συχνά ατόμων που έχουν λάβει συστάσεις στο παρελθόν – ο recommender συστήνει στοιχεία που πιθανώς θα επιθυμούσε

αυτός που ζητάει τη σύσταση. Επιπλέον, ο recommender ενδέχεται να προσδιορίσει άτομα με παρόμοια ενδιαφέροντα. Ο αιτών τη σύσταση μπορεί να χρησιμοποιήσει τη σύσταση για να επιλέξει στοιχεία από το σύμπαν ή να επικοινωνήσει με άλλους ομοϊδεάτες.



Σχήμα 1: Recommendation Process Μοντέλο (Terveen L. και Hill W., 2001)

Το μοντέλο αυτό προορίζεται να είναι γενικό και να καλύπτει ένα ευρύ φάσμα δραστηριοτήτων σύστασης.

2.1.2 Θέματα για υπολογιστικά Συστήματα Σύστασης

Ένα υπολογιστικό σύστημα σύστασης αυτοματοποιεί ή υποστηρίζει μέρος της διαδικασίας σύστασης. Ένα αυτόματο σύστημα σύστασης προϋποθέτει τον ρόλο του recommender: προσφέρει συστάσεις στους χρήστες με βάση τις προτιμήσεις τους (και ίσως επίσης με βάση τις προτιμήσεις άλλων ατόμων). Ένα σύστημα υποστηρίξης συστάσεων διευκολύνει τη δημιουργία και την κοινή χρήση των συστάσεων. Μπορούμε να προσδιορίσουμε τέσσερα κύρια θέματα για το χαρακτηρισμό του χώρου σχεδίασης για Συστήματα Σύστασης. Τα ζητήματα αφορούν **προτιμήσεις, ρόλους και επικοινωνία, αλγόριθμους, και αλληλεπίδραση ανθρώπου-υπολογιστή**. Θα παρουσιαστεί το κάθε ένα με συντομία παρακάτω.

Προτιμήσεις

Η σύσταση βασίζεται στις προτιμήσεις. Έτσι, ένα αυτοματοποιημένο σύστημα σύστασης πρέπει να αποκτήσει προτιμήσεις από άτομα σχετικά με τον αντίστοιχο τομέα. Αυτό θέτει ορισμένα ερωτήματα, συμπεριλαμβανομένων των εξής:

- Ποιανού οι προτιμήσεις χρησιμοποιούνται; Αυτές του προσώπου που ζητάει τη σύσταση, αυτές των προηγούμενων χρηστών του συστήματος; Ή μήπως προτιμήσεις εκφρασμένες από άτομα σε ένα εντελώς διαφορετικό πλαίσιο, όπως ένα δημόσιο φόρουμ (π.χ. chat room, bulletin board, ή newsgroup);
- Με ποιόν τρόπο λαμβάνονται οι προτιμήσεις; Για παράδειγμα, οι χρήστες πρέπει να εκφράσουν τις δικές τους προτιμήσεις ως μέρος της διαδικασίας της εξεύρεσης μιας σύστασης; Οι προτιμήσεις εκφράζονται άμεσα ή έμμεσα;
- Τι κίνητρα υπάρχουν για τα άτομα να προσφέρουν τις προτιμήσεις τους;
- Ποιά είναι η μορφή μιας προτίμησης; Πώς απεικονίζονται οι προτιμήσεις;

Οι ρόλοι και η επικοινωνία

- Ο ρόλος του recommender υποστηρίζεται από ένα υπολογιστικό σύστημα ή από ένα άτομο; Εάν ισχύει το δεύτερο, τι ρόλο παίζουν οι υπολογισμοί στην υποστήριξη της διαδικασίας σύστασης;
- Οι άνθρωποι παίζουν διακριτούς ρόλους ή όλοι οι χρήστες του συστήματος παίζουν τον ίδιο ρόλο; Οι ρόλοι είναι καθορισμένοι, ή εξελίσσονται;
- Πώς ξεκινάει η αλληλεπίδραση ανάμεσα στον χρήστη του συστήματος σύστασης και τον recommender; Ποιος την εκκινεί; Η σύσταση κατευθύνεται σε ένα συγκεκριμένο πρόσωπο ή

μεταδίδεται σε οποιονδήποτε ενδιαφέρεται; Υπάρχει η δυνατότητα για τους χρήστες του συστήματος να επανατροφοδοτήσουν με τα σχόλιά τους τον recommender;

- Ποιές πληροφορίες που αφορούν τα άτομα των οποίων οι προτιμήσεις χρησιμοποιούνται στον υπολογισμό μιας σύστασης, αποκαλύπτονται στο χρήστη του συστήματος σύστασης; Υπάρχει ευκαιρία για τη διαμόρφωση κοινοτήτων ομοϊδεατών; Εάν αποκαλύπτεται πληροφορία σχετικά με αυτούς που παρέχουν τις προτιμήσεις τους, λαμβάνονται μέτρα διασφάλισης του ιδιωτικού απόρρητου;

Αλγόριθμοι για υπολογισμό των συστάσεων

- Με ποιόν τρόπο ένα σύστημα αυτόματης σύστασης καθορίζει, ποιανού τις προτιμήσεις να χρησιμοποιήσει στον υπολογισμό της σύστασης; Εάν θεωρήσουμε όλα τα άτομα που έχουν εκφράσει τις προτιμήσεις τους για έναν δεδομένο τομέα τοποθετημένα σε ένα μεγάλο, πολυδιάστατο χώρο, αυτό είναι το πρόβλημα της εξεύρεσης γειτόνων σε αυτόν το χώρο για το πρόσωπο που ζητά μια σύσταση.
- Με ποιόν τρόπο υπολογίζονται οι συστάσεις; Για παράδειγμα, δεδομένου ότι έχει οριστεί ένα σύνολο γειτόνων για αυτόν που ζητά τη σύσταση, πώς σταθμίζονται και συνδυάζονται οι προτιμήσεις αυτών των γειτόνων;

Αλληλεπίδραση ανθρώπου-υπολογιστή

- Με ποιόν τρόπο παρουσιάζονται οι συστάσεις στο πρόσωπο το οποίο τις ζήτησε; Το πιο απλό και κοινό παράδειγμα είναι μια ταξινομημένη λίστα. Πιο περίπλοκα παραδείγματα περιλαμβάνουν 2D και 3D απεικονίσεις, καθώς και οπτικά σχόλια των υπαρχουσών διαστημάτων πληροφορίας.

2.2 Τύποι Συστημάτων Σύστασης

Πολλά διαφορετικά συστήματα σύστασης αναπτύχθηκαν κατά τη διάρκεια της δεκαετίας του 1990 και μαζί μ'αυτά αναπτύχθηκε και η ορολογία με όρους όπως "συνεργατική διοίκηση" (Collaborative filtering) και "κοινωνική δικτύωση" (social networking). Θα αναφέρουμε τις διαφορετικές προσεγγίσεις όσον αφορά στα τέσσερα ζητήματά που έχουν αναφερθεί παραπάνω. Οι προσεγγίσεις αυτές μπορούν να διακριθούν ανάλογα με το σε ποιο από τα τέσσερα βασικά ζητήματα επικεντρώνονται, και με ποιόν τρόπο τα αντιμετωπίζουν.

- Τα Συστήματα Σύστασης με βάση το περιεχόμενο (content-based Recommender Systems) χρησιμοποιούν μόνο τις προτιμήσεις του χρήστη, επιχειρούν να προτείνουν αντικείμενα που είναι παρόμοια με αυτά που άρεσαν στο χρήστη στο παρελθόν. Επικεντρώνονται σε αλγόριθμους για μάθηση των προτιμήσεων του χρήστη και στο φιλτράρισμα νέων αντικειμένων που να ταιριάζουν περισσότερο με τις προτιμήσεις του χρήστη.
- Τα Συστήματα Υποστήριξης Σύστασης (Recommendation support systems) δεν αυτοματοποιούν τη διαδικασία σύστασης, ως εκ τούτου, δεν πρέπει να παρουσιάσουν προτιμήσεις ή να υπολογίσουν συστάσεις. Αντίθετα, χρησιμεύουν ως εργαλεία για την υποστήριξη ατόμων στην κοινή χρήση συστάσεων, βοηθώντας τόσο αυτούς που παράγουν συστάσεις, όσο και αυτούς που αναζητούν μια σύσταση.
- Τα Συστήματα Εξόρυξης Κοινωνικών Δεδομένων (Social data mining systems) εξορύσσουν τις έμμεσες προτιμήσεις από υπολογιστικές εγγραφές κοινωνικής δραστηριότητας, όπως μηνύματα Usenet, ιστορικό χρήσης του συστήματος, αναφορές ή υπερ-συνδέσεις. Αυτά τα συστήματα επίσης επικεντρώνονται σε ζητήματα αλληλεπίδρασης Ανθρώπου-Υπολογιστή (Human computer Interaction), που αφορούν σε οπτικοποίηση των αποτελεσμάτων. Αυτές οι παρουσιάσεις συχνά έχουν παρουσιαστεί για να βοηθήσουν την πλοήγηση των διαστημάτων πληροφορίας όπως το World Wide Web και αυτό βοήθησε στην καθιέρωση του όρου social navigation (κοινωνική πλοήγηση).

- Τα συστήματα Συνεργατικής διοίκησης (Collaborative filtering systems) απαιτούν από τους χρήστες να εκφράσουν προτιμήσεις αξιολογώντας μια δωδεκάδα ή δύο αντικείμενα, συγχωνεύοντας τους ρόλους αυτού που ζητάει τη σύσταση και αυτού που παρέχει την προτίμηση. Αυτά τα συστήματα επικεντρώνονται σε αλγόριθμους για το ταίριασμα ατόμων με βάση τις προτιμήσεις τους και τη στάθμιση των ενδιαφερόντων των ατόμων με παρόμοια γούστα για να παράξουν μια σύσταση για τον χρήστη που τη ζητάει.

<i>Issues</i>	<i>Approaches</i>			
	Content-based	Rec. Support	Social Data Mining	Collaborative Filtering
Preferences	Seeker's preferences only		Mines preferences; seeker's preferences typically not used	Seekers must state preferences
Roles & Communication	System automates Role asymmetry	System supports human recommenders and seekers	System automates High potential for community; raises significant privacy concerns Role asymmetry vs. role uniformity	
Algorithms	Machine learning, information retrieval		Data mining	Preference matching and weighting
HCI			Visualization; visual annotation	

Πίνακας 1: Συστήματα σύστασης θέματα και προσεγγίσεις (Terveen L. και Hill W., 2001)

Ο Πίνακας 1 συνοψίζει τις τέσσερις κύριες προσεγγίσεις και τα θέματα στα οποία επικεντρώνονται. Θεωρούμε τις τέσσερις κύριες προσεγγίσεις με περισσότερες λεπτομέρειες στη συνέχεια του κεφαλαίου. Θα περιγράψουμε τους διαφορετικούς τρόπους που υποστηρίζουν τη διαδικασία σύστασης και θα προσδιορίσουμε τις προκλήσεις και τις ευκαιρίες για μελλοντικές εργασίες.

2.2.1 Συστήματα Σύστασης με βάση το περιεχόμενο

Τα Συστήματα Σύστασης με βάση το περιεχόμενο (Content-based recommenders) (Lieberman 1997, Maes 1994) βασίζονται στην ιδέα "βρες μου πράγματα όπως αυτά που μου έχουν αρέσει στο παρελθόν". Μαθαίνουν προτιμήσεις μέσω της επανατροφοδότησης από τον χρήστη. Η επανατροφοδότηση μπορεί να είναι άμεση (explicit) – για παράδειγμα, οι χρήστες μπορούν να αξιολογήσουν πράγματα ως "καλά" ή "κακά". Η η επανατροφοδότηση μπορεί να είναι έμμεση (implicit) – για παράδειγμα, με βάση το αν οι χρήστες επιλέγουν να διαβάσουν ένα προτεινόμενο έγγραφο και πόσο χρόνο δαπανούν στο διάβασμα. Οι προτιμήσεις παρουσιάζονται ως ένα προφίλ των ενδιαφερόντων των χρηστών σε συγκεκριμένους τύπους περιεχομένου, συχνά εκφρασμένες ως ένα σύνολο σταθμισμένων λέξεων-κλειδιά. Τεχνικές από τη μηχανική μάθηση και την ανάκτηση πληροφοριών εφαρμόζονται για την εκμάθηση και την παρουσίαση των προτιμήσεων του χρήστη.

Τα Content-based συστήματα σύστασης δεν είναι αυτά που θα μας απασχολήσουν εδώ. Ωστόσο, αποτελούν ένα σημείο αντίθεσης που βοηθάει στο να διευκρινιστεί ο τύπος των κοινωνικών συστημάτων σύστασης (Social Recommender Systems). Τα κοινωνικά συστήματα σύστασης δημιουργούν μια Συστήματα Σύστασης βασισμένα σε Τεχνικές Νοημοσύνης Σμήνους

ενδιάμεση (ίσως έμμεση) αλληλεπίδραση μεταξύ κάποιου που επιδιώκει μια σύσταση και ενός συνόλου ατόμων που έχουν εκδηλώσει προηγουμένως σχετικές προτιμήσεις.

Τα Content-based και τα social συστήματα σύστασης, έχουν συμπληρωματικά πλεονεκτήματα. Για παράδειγμα, εάν, στο παρελθόν, κάποιος προτιμούσε βιβλία σχετικά με την εξερεύνηση του Άρη, είναι πιθανό να ενδιαφέρεται για ένα νέο βιβλίο σχετικά με τον Άρη, ανεξάρτητα από μια σύσταση από οποιονδήποτε άλλο. (Με άλλα λόγια, αυτό το βιβλίο προτείνεται με βάση το περιεχόμενό του.) Από την άλλη, ένας φίλος μπορεί να συστήσει ένα βιβλίο σχετικά με ένα εντελώς νέο θέμα, π.χ. σχετικά με το ρόλο των ασθενειών στην έκβαση του αποτελέσματος των μαχών στην ιστορία. Εάν δεχτούμε τη σύσταση και το βιβλίο μας αρέσει, μπορούμε να αναπτύξουμε ένα εντελώς νέο ενδιαφέρον. Αυτή η δυναμική της τυχαίας ανακάλυψης είναι πολύ σημαντική, αφού μπορεί να βοηθήσει τα άτομα να αλλάξουν τη ρουτίνα τους και να διευρύνουν τους ορίζοντές τους.

Υπάρχει επίσης μια καθοριστική διαφορά από την άποψη των συστημάτων, τα με βάση το περιεχόμενο συστήματα σύστασης πρέπει να είναι σε θέση να παρουσιάζουν και να χειρίζονται το περιεχόμενο των στοιχείων. Αυτό είναι τεχνικά δύσκολο ακόμη και στην πιο καλά κατανοητή υπόθεση, δηλαδή, για κείμενο, και σήμερα είναι σχεδόν αδύνατο για τα είδη που δεν είναι κείμενο, όπως η μουσική ή οι εικόνες. Ωστόσο, οι κοινωνικές προσεγγίσεις δεν έχουν κανένα τέτοιο πρόβλημα, εφόσον δεν (πρέπει να) επεξεργάζονται το περιεχόμενο. Αντίθετα, λειτουργούν με τις προτιμήσεις του χρήστη, τις γνώμες και τις συμπεριφορές. Λόγω των συμπληρωματικών πτυχών των με βάση το περιεχόμενο και των κοινωνικών recommenders, μια ελκυστική τακτική έρευνας είναι η δημιουργία υβριδικών συστημάτων.

2.2.2 Συστήματα Υποστήριξης Συστάσεων

Τα **Συστήματα Υποστήριξης Συστάσεων (Recommendation Support Systems)** είναι υπολογιστικά εργαλεία για την υποστήριξη των ατόμων στη φυσική δραστηριότητα της ανταλλαγής συστάσεων, συμπεριλαμβανομένων τόσο της δημιουργίας, όσο και της εύρεσής τους.

Μια ομάδα ερευνητών στη Xerox PARC ανέπτυξαν το Tapestry, το πρώτο σύστημα υποστήριξης συστάσεων (Goldberg et al 1992). Το Tapestry ήταν ένα ηλεκτρονικό σύστημα ανταλλαγής μηνυμάτων που επέτρεπε στους χρήστες είτε να βαθμολογούν τα μηνύματα (σαν "καλά" ή "κακά"), είτε να συσχετίζουν σχόλια ελεύθερου κειμένου με τα μηνύματα (το Tapestry συνήθως θεωρείται το πρώτο σύστημα σύστασης οποιουδήποτε είδους).

Τα μηνύματα ήταν αποθηκευμένα σε μια βάση δεδομένων και μπορούσαν να ανακτηθούν με βάση όχι μόνο το περιεχόμενό τους, αλλά και με βάση τις γνώμες των άλλων. Για παράδειγμα, κάποιος μπορούσε να ανακτήσει έγγραφα που είχαν πάρει υψηλή βαθμολογία από ένα συγκεκριμένο άτομο ή άτομα, ή μπορούσε να ανακτήσει έγγραφα των οποίων τα σχόλια περιείχαν συγκεκριμένες λέξεις-κλειδιά.

Οι Maltz και Ehrlich (1995) ανέπτυξαν περαιτέρω αυτή την προσέγγιση. Παρατήρησαν την υφιστάμενη πρακτική μέσα στις εταιρίες και πρόσεξαν ότι ορισμένα άτομα παίζουν πάντα έναν πολύ δραστήριο ρόλο στη διατύπωση συστάσεων. Δημιούργησαν ένα σύστημα που έχει σχεδιαστεί ρητά για την υποστήριξη των δύο διακριτών ρόλων της σύστασης: του δημιουργού της σύστασης και του χρήστη της σύστασης (του αιτούντα τη σύσταση). Το σύστημά τους έδωσε τη δυνατότητα στα άτομα να δημιουργήσουν συστάσεις που αποτελούνταν από δείκτες σε έγγραφα, τα οποία θα μπορούσε να είναι οργανωμένα σε σύντομες ανασκοπήσεις (digests). Οι συστάσεις στη συνέχεια θα μπορούσαν να απευθυνθούν σε καθορισμένους συναδέλφους. Το σύστημα υποστήριζε επίσης τους χρήστες στο διάβασμα αυτών των digests.

Παρόμοιες ιδέες είναι δημοφιλείς στο World Wide Web από την δημιουργία του μέχρι και στις μέρες μας με τη μορφή των weblogs (ιστολόγιων) που έχουν κερδίσει την προσοχή του κοινού (Barrett 1999, Katz 1999). Ολοένα και περισσότεροι άνθρωποι δημιουργούν σχολιασμένα αρχεία καταγραφής των συνδέσεων. Επιλέγουν συνδέσμους και γράφουν σχόλια προκειμένου να αντικατοπτρίσουν τα ιδιαίτερα τους ενδιαφέροντα και τις αντιλήψεις. Πολλά ιστολόγια ενημερώνονται καθημερινά, έτσι, το

περιεχόμενό τους είναι φυσικά διατεταγμένο χρονολογικά, αλλά ορισμένα προσφέρουν επίσης επίκαιρη κατηγοριοποίηση των συνδέσεων.

Ορισμένα ιστολόγια γίνονται εξ ολοκλήρου από ένα άτομο, ορισμένα είναι αποτέλεσμα ομαδικής προσπάθειας, και ορισμένα εμπíπτουν κάπου ενδιάμεσα. Κάποια έχουν χιλιάδες αναγνώστες, ενώ άλλα έχουν μόνο μερικούς. Μερικά εστιάζονται έντονα σε ένα μόνο θέμα (π.χ., νέα μέσα μαζικής ενημέρωσης), αλλά τα περισσότερα τείνουν να είναι αρκετά εκλεκτικά (δεν αποτελεί έκπληξη το γεγονός, εφόσον έχουν την τάση να περικλείουν ό,τι οι συντάκτες τους βρίσκουν ενδιαφέρον και οι περισσότεροι άνθρωποι έχουν περισσότερα από ένα ενδιαφέροντα).

Το About.com εμπορευματοποίησε μια σχετική έννοια, αυτή του ανθρώπινου οδηγού θεμάτων. Η τοποθεσία About.com φιλοξενεί εκατοντάδες πεδία συγκεκριμένων θεμάτων, κάθε ένα από τα οποία συντηρείται από έναν εμπειρογνώμονα στο συγκεκριμένο θέμα. Το περιεχόμενο σε κάθε θεματικό πεδίο περιλαμβάνει οργανωμένες και σχολιασμένες συλλογές web συνδέσεων, ειδήσεις, δημοσκοπήσεις, συχνές ερωτήσεις και άλλα χαρακτηριστικά "κοινότητας", καθώς και εμπορικά χαρακτηριστικά.

Τα Συστήματα υποστήριξης συστάσεων ακολουθούν την υφιστάμενη πρακτική αρκετά στενά. Δεν προϋποθέτουν νέους ρόλους ή νέες δραστηριότητες. Αντίθετα, στηρίζονται σε ένα καλά βεβαιωμένο, φυσικό καταμερισμό της εργασίας: μερικοί άνθρωποι είναι ιδιαίτερα πρόθυμοι να παράγουν συστάσεις, ενώ οι περισσότεροι, τις περισσότερες φορές, προτιμούν να τις χρησιμοποιούν. Τα Συστήματα υποστήριξης συστάσεων είναι αποτελεσματικά, όταν υπάρχουν αρκετά άτομα διατεθειμένα να βρουν και να συστήσουν πληροφορίες. Σχεδόν ποτέ τα άτομα δεν πληρώνονται για να το κάνουν αυτό, συνήθως, είναι εργασία που προσφέρεται από αγάπη. Επιπλέον, οι ανάγκες των δύο πλευρών, αυτών που κάνουν τις συστάσεις και των χρηστών των συστάσεων, πρέπει να πληρούνται. Για αυτούς που κάνουν τις συστάσεις (recommenders), οι κύριες ανάγκες είναι η αναγνώριση και η επανατροφοδότηση, η πρόκληση ενθαρρυντικών, προκλητικών ή γεμάτων ευγνωμοσύνη αποκρίσεων είναι συχνά αυτό και μόνο που θέλουν οι recommenders. Για τους χρήστες, οι συστάσεις που λαμβάνουν πρέπει να είναι σχετικές και ενδιαφέρουσες. Το να πλημμυρίζει το γραμματοκιβώτιό μας με μηνύματα που ένας φίλος βρίσκει ενδιαφέροντα, αλλά εμείς όχι, είναι απλώς μια εξατομικευμένη εκδοχή της ανεπιθύμητης αλληλογραφίας (spam).

Οι συντάκτες των weblogs που ασχολούνται με την «επιχείρηση των συστάσεων» έχουν πολλές ανάγκες. Πρώτον, το να διατηρήσουν τα ιστολόγια τους χρήσιμα κατά τη διάρκεια των ετών τους φέρνει αντιμέτωπους με ένα σημαντικό έργο διαχείρισης των πληροφοριών. Χρειάζεται να ελέγχουν τις παλιές συνδέσεις. Χρειάζεται επίσης να παρέχουν μια μη-χρονολογική οργάνωση του περιεχομένου, δηλαδή ανάπτυξη κατηγοριών περιεχομένου στις οποίες μπορεί να τοποθετηθούν οι συνδέσεις, ή ευρετηρίαση του περιεχομένου και δημιουργία μιας μηχανής αναζήτησης. Δεύτερον, μπορεί να χρειάζονται συστήματα σύστασης που να προτείνουν νέα και ενδιαφέροντα στοιχεία, τα οποία να προσαρμόζονται στο θέμα τους. Τέλος, ζωτικής σημασίας είναι οι τεχνικές που βοηθούν τους recommenders να βρουν το κατάλληλο κοινό. Η αύξηση των απόψεων, όπως παρουσιάζεται με τον αυξανόμενο αριθμό των ιστολογίων, φαίνεται σχεδόν να εγγυάται ότι ο κάθε recommender μπορεί να βρει ένα κοινό, και ότι ο καθένας που αναζητά πληροφορίες μπορεί να βρει έναν ομοϊδεάτη οδηγό. Ωστόσο, το οικείο φάντασμα της υπερφόρτωσης της πληροφορίας εμφανίζεται σύντομα: με ολοένα περισσότερες επιλογές, πώς μπορούν οι άνθρωποι να εντοπίσουν ο ένας τον άλλο;

2.2.3 Social data mining systems

Ο καθένας θα μπορούσε να περπατήσει στα χνάρια των ειδικών. Θα μπορούσαμε να χρησιμοποιήσουμε τη μεταφορά ενός μονοπατιού στο δάσος. Ωστόσο, αυτή η μεταφορά υπογραμμίζει τον ρόλο της συλλογικής προσπάθειας, περισσότερο από αυτή του ατόμου. Ένα μονοπάτι είναι το αποτέλεσμα των αποφάσεων πολλών ατόμων, ενωμένων μόνο από το πού θα επιλέξουν να βαδίσουν, και αντικατοπτρίζει

τι θεωρούν οι πεζοπόροι ένα καλό μονοπάτι. Το μονοπάτι αντανακλά το ιστορικό της χρήσης και χρησιμεύει ως πόρος για μελλοντικούς χρήστες.

Οι προσεγγίσεις της εξόρυξης κοινωνικών δεδομένων (Social data mining) αναζητούν ανάλογες καταστάσεις στον κόσμο της υπολογιστικής. Οι ερευνητές αναζητούν περιπτώσεις κατά τις οποίες ομάδες ατόμων παράγουν υπολογιστικές εγγραφές (όπως έγγραφα, μηνύματα Usenet ή τοποθεσίες web και συνδέσμους) σαν μέρος της κανονικής τους δραστηριότητας. Δυνητικά χρήσιμες πληροφορίες, έμμεσες σ' αυτές τις εγγραφές, προσδιορίζονται, υπολογιστικές τεχνικές για τη συγκέντρωση της πληροφορίας επινοούνται, και σχεδιάζονται τεχνικές οπτικοποίησης για την παρουσίαση των αποτελεσμάτων. Έτσι, οι υπολογισμοί ανακαλύπτουν και καθιστούν σαφή τα "μονοπάτια μέσα από το δάσος" που δημιουργήθηκαν από κοινότητες συγκεκριμένων χρηστών.

Ο παγκόσμιος Ιστός, με το εμπλουτισμένο περιεχόμενο, τη δομή των συνδέσεων, και τη χρήση των αρχείων καταγραφής, υπήρξε ένας σημαντικός τομέας για την έρευνα της εξόρυξης κοινωνικών δεδομένων. Μια βασική αντίληψη είναι ότι, μια σύνδεση από μια τοποθεσία Web σε μια άλλη συχνά υποδεικνύει ομοιότητα του περιεχομένου μεταξύ των δύο τοποθεσιών. Διάφοροι αλγόριθμοι συσταδοποίησης και αξιολόγησης έχουν σχεδιαστεί για να τυποποιήσουν αυτή την αντίληψη. Ο αλγόριθμος του Kleinberg (1998) είναι ένα πολύ γνωστό παράδειγμα. Στον εμπορικό κόσμο, η μηχανή αναζήτησης Google (www.google.com) χρησιμοποιεί έναν παρόμοιο αλγόριθμο ανάλυσης συνδέσμων για την ομαδοποίηση και ταξινόμηση των URLs. Άλλες εργασίες επικεντρώθηκαν στην εξαγωγή πληροφοριών από αρχεία καταγραφής της χρήσης του web. Το σύστημα Footprints καταγράφει το ιστορικό πλοήγησης του χρήστη (Wexelblat και Maes P. 1990), το αναλύει για να βρει τους κοινούς συνδέσμους μεταξύ των web σελίδων και κατασκευάζει πολλές διαφορετικές απεικονίσεις αυτών των δεδομένων για να βοηθήσει την πλοήγηση του χρήστη μέσα σε μια τοποθεσία web.

Ο Chalmers και οι συνεργάτες του (Chalmers et al 1998) θεώρησαν το μονοπάτι δραστηριότητας (activity path) – π.χ., μια ακολουθία των διευθύνσεων URLs που επισκέφθηκε κάποιος κατά τη διάρκεια μιας περιόδου περιήγησης – ως τη βασική μονάδα. Ανέπτυξαν τεχνικές για τον υπολογισμό ομοιοτήτων μεταξύ των μονοπατιών για να προβούν σε συστάσεις σ' αυτή τη βάση – για παράδειγμα, να συστήσει σελίδες σε κάποιον που κάποιος άλλος περιηγήθηκε και έχουν στενή εγγύτητα με τις σελίδες που αυτός έχει περιηγηθεί. Άλλες τεχνικές εξάγουν πληροφορίες από πολλές πηγές. Για παράδειγμα, οι Pirolli, Pitkow και Rao (Pirolli 1996, Pitkow 1997) συνδύασαν web συνδέσεις με τα δεδομένα χρήσης του web και την ομοιότητα του κειμένου για την κατηγοριοποίηση και συσταδοποίηση των ιστοσελίδων.

Άλλες εργασίες έχουν επικεντρωθεί στην εξαγωγή πληροφοριών από online συνομιλίες, όπως το Usenet. Το PHOAKS (Hill 1992) εξορύσσει μηνύματα σε ομάδες συζήτησης του Usenet (Usenet newsgroups) αναζητώντας αναφορές σε σελίδες Web. Κατηγοριοποιεί και συγκεντρώνει αναφορές για τη δημιουργία λιστών δημοφιλών ιστοσελίδων για κάθε ομάδα. Ο Donath και οι συνεργάτες του (Viegas and Donath 1990) έχουν συλλέξει πληροφορίες από ομάδες συζήτησης του Usenet και συνομιλίες και τις χρησιμοποίησαν για τη δημιουργία απεικονίσεων της συνομιλίας. Αυτές οι απεικονίσεις μπορούν να χρησιμοποιηθούν για την εύρεση συζητήσεων με επιθυμητές ιδιότητες, όπως η ισότητα συμμετοχής ή οι πολλοί τακτικοί συμμετέχοντες.

Τα Συστήματα εξόρυξης κοινωνικών δεδομένων δεν απαιτούν από τους χρήστες να συμμετάσχουν σε οποιαδήποτε νέα δραστηριότητα, μάλλον, επιδιώκουν να εκμεταλλευτούν πληροφορίες σχετικά με τις προτιμήσεις του χρήστη, έμμεσες στις εγγραφές της υπάρχουσας δραστηριότητας.

Όπως και τα συστήματα υποστήριξης συστάσεων, εργάζονται σε καταστάσεις όπου τα άτομα έχουν διαφορετικούς ρόλους, δηλαδή, λίγοι παράγουν και μοιράζονται απόψεις και προτιμήσεις, ενώ τα περισσότερα άτομα είναι ευχαριστημένα χρησιμοποιώντας αυτές τις πληροφορίες, όταν τις έχουν ανάγκη.

Τα Συστήματα μπορούν να διατηρούν και να διαβιβάζουν πληροφορίες σχετικά με το πλαίσιο της δραστηριότητας από την οποία έχουν εξαχθεί οι προτιμήσεις. Αυτό μπορεί να οδηγήσει σε πιο

ενημερωτικές και πιθανώς περισσότερο χρήσιμες συστάσεις. Επίσης, δημιουργεί ευκαιρίες για την οικοδόμηση μιας κοινότητας – οι άνθρωποι μπορούν να έρθουν σε επαφή με άλλους που συμμερίζονται τις προτιμήσεις τους. Ωστόσο, σε αντίθεση με τα συστήματα υποστήριξης συστάσεων, που βοηθούν τα άτομα που προτίθενται να μοιραστούν τις συστάσεις τους, τα συστήματα εξόρυξης κοινωνικών δεδομένων εξάγουν δεδομένα προτίμησης από περιβάλλοντα όπου οι παροχές των δεδομένων μπορεί να μην είχαν καμία τέτοια πρόθεση. Έτσι, οι δυνατότητες για την οικοδόμηση κοινότητας πρέπει να είναι σε ισορροπία σε σχέση με την ιδιωτικότητα των ατόμων που παρήγαγαν κατά κύριο λόγο τις προτιμήσεις.

Τα περισσότερα συστήματα εξόρυξης κοινωνικών δεδομένων, δημιουργούν "broadcast" συστάσεις, δηλαδή, οι συστάσεις είναι διαθέσιμες (ίσως ως απεικονίσεις και βοηθήματα περιήγησης) σε όποιον χρησιμοποιεί το σύστημα. Ωστόσο, αυτό δεν είναι υποχρεωτικό: εάν οι προτιμήσεις εξαχθούν και συσχετισθούν με τα άτομα που τις έχουν παράξει, ένας αλγόριθμος μπορεί να ταιριάζει χρήστες με βάση τις προτιμήσεις τους και, ως εκ τούτου, να υπολογίσει προσωπικές συστάσεις. Το σύστημα των Chalmers et al (1998) είναι ένα παράδειγμα συστήματος που το κάνει αυτό.

Το πρώτο σύνολο ζητημάτων αφορά τα δεδομένα που εξορύσσονται και τους αλγόριθμους εξόρυξης. Η εμπειρία που έχει αποκτηθεί από τα συστήματα PHOAKS και TopicShop χρησιμοποιήθηκε για την εξήγηση αυτών των ζητημάτων.

- Υπάρχουν χρήσιμα δεδομένα (δηλαδή προτιμήσεις) κρυμμένα στις εγγραφές δραστηριότητας; Πειράματα που έγιναν στα πλαίσια του project PHOAKS το 1996 έδειξαν ότι σχεδόν το ένα τέταρτο όλων των μηνυμάτων του Usenet περιείχαν αναφορές σε διευθύνσεις URLs και περίπου το 20% του χρόνου τα άτομα που ανέφεραν μια διεύθυνση URL, εξέφραζαν μια προτίμηση για αυτή. Αυτό σημαίνει ότι υπάρχουν πολλές χιλιάδες URL συστάσεις στο Usenet κάθε μέρα. Ως εκ τούτου, η επόμενη πρόκληση είναι:
- Μπορούν τα δεδομένα να εξαχθούν με ακρίβεια και αποτελεσματικότητα? Τα πειράματα του PHOAKS έδειξαν ότι οι κανόνες για την κατάταξη των αναφορών των URLs ως συστάσεων ήταν σχεδόν στο 90% σωστοί (και στην ακρίβεια και στην ανάκληση).
- Είναι τα εξαγόμενα δεδομένα υψηλής ποιότητας? Έπρεπε να διερευνηθεί αν οι διευθύνσεις URLs που συνιστούσε το PHOAKS για ένα συγκεκριμένο θέμα ήταν πραγματικά καλές συστάσεις. Συγκεκριμένα, ερευνήθηκε το αν ο τρόπος μέτρησης του PHOAKS – που αντιστοιχεί μία ψήφο σε ένα URL για κάθε ξεχωριστό πρόσωπο το οποίο το συνιστά – συνάδει με την ανθρώπινη κρίση περί ποιότητας. Αποδείχτηκε μια θετική συσχέτιση μεταξύ αυτού του μέτρου και της πιθανότητας μια συγκεκριμένη διεύθυνση URL να περιλαμβάνεται σε μια λίστα FAQ (Frequently Asked Question list) για την σχετική ομάδα συζήτησης. Με άλλα λόγια, όσο περισσότερα τα άτομα που συνιστούν ένα URL, τόσο πιθανότερο ήταν να το έχει συμπεριλάβει σε μια λίστα σχετικών πηγών ένας μεμονωμένος εμπειρογνώμονας επί του θέματος.

Εξετάστηκε επίσης αυτό το ερώτημα για το σύστημα TopicShop (Amento 1990). Το TopicShop εξορύσσει πληροφορίες από σελίδες web και συνδέσεις και η διεπαφή του παρέχει στους χρήστες εύκολη πρόσβαση σε αυτήν τη διεπαφή. Πειράματα έδειξαν ότι το TopicShop ήταν σε θέση να επιλέγει περισσότερο υψηλής ποιότητας τοποθεσίες κατά 80%, ενώ χρειαζόνταν λιγότερο χρόνο και εξετάζοντας λιγότερες τοποθεσίες από τους χρήστες που δεν είχαν πρόσβαση σε αυτά τα δεδομένα.

Άλλα θέματα προκύπτουν όταν αυτοί που ζητούν τη σύσταση ενδιαφέρονται το ίδιο ή και παραπάνω για την εύρεση ενός ατόμου, όπως και για την εύρεση πληροφοριών. Το ReferralWeb (Kautz et al 1997, Terveen 1997) αναλύει έγγραφα web, επεξεργάζεται την εμφάνιση ονομάτων μέσα σε έγγραφα για να δημιουργήσει ένα κοινωνικό δίκτυο, και συσχετίζει άτομα με τις ειδικές γνώσεις τους.

Αυτό το ζήτημα συνδέεται με το θέμα της συνειδητοποίησης (awareness) στα συνεργατικά συστήματα (Dourish 1992). Τα συστήματα σύστασης μπορούν να αντιμετωπίσουν το ζήτημα του ποιός πρέπει να ενημερωθεί για ποιούς, δηλαδή, πώς σχηματίζονται οι κοινότητες. Οι Maglio et al (Maglio 2000) προσδιόρισαν έναν αριθμό διαφορετικών τεχνικών για τον καθορισμό των κοινοτήτων στο web. Οι κοινότητες μπορούν να σχηματιστούν από άτομα από την ίδια εταιρεία, από χρήστες που περιηγούνται

στις ίδιες ή σε στενά συνδεδεμένες σελίδες, ή από χρήστες που έχουν παρόμοιο ιστορικό περιήγησης (Chalmers 1998).

2.2.4 Συστήματα Συνεργατικής Διοίκησης

Θα περίμενε κανείς να λάβει την καλύτερη σύσταση/πρόταση από κάποιον με παρόμοιο γούστο. Το πρόβλημα, ωστόσο, είναι πώς να βρεθεί αυτός. Ίσως χρειαστεί να αλληλεπιδράσετε με πολλά διαφορετικά άτομα, και να μάθετε σιγά-σιγά τις προτιμήσεις τους, πριν να ξεκινήσετε να λαμβάνετε συστάσεις, τις οποίες να εμπιστευτείτε.

Η Συνεργατική Διοίκηση διερευνά την τεχνική του ταιριάσματος ατόμων με παρόμοια ενδιαφέροντα και στη συνέχεια τη διατύπωση συστάσεων σ' αυτή τη βάση. Τρεις πυλώνες της παρούσας προσέγγισης είναι (1) πρέπει να συμμετέχουν πολλά άτομα (έτσι ώστε να είναι πιθανό ότι ο καθένας θα μπορεί να βρει άλλα άτομα με παρόμοιες προτιμήσεις), (2) πρέπει να υπάρχει ένας εύκολος τρόπος για τα άτομα να παρουσιάζουν τα ενδιαφέροντά τους στο σύστημα, και (3) οι αλγόριθμοι θα πρέπει να είναι ικανοί να ταιριάζουν άτομα με παρόμοια ενδιαφέροντα.

Η Συνεργατική Διοίκηση έκανε τη δουλειά του χρήστη πολύ απλή: εκφράζει τις προτιμήσεις του βαθμολογώντας εμπορεύματα (όπως βιβλία, ταινίες ή CD) που του παρουσιάζει το σύστημα. Αυτή η βαθμολόγηση θα αποτελέσει στη συνέχεια μια κατά προσέγγιση εκπροσώπηση των προτιμήσεών του σε αυτόν τον τομέα. Το σύστημα, στη συνέχεια, αντιστοιχίζει αυτές τις βαθμολογήσεις με τις βαθμολογήσεις που υποβάλλονται από όλους τους άλλους χρήστες του συστήματος. Το αποτέλεσμα είναι το σύνολο των "πλησιέστερων γειτόνων" προς τον χρήστη, κι αυτό τυποποιεί την έννοια των ατόμων με παρόμοια γούστα. Τέλος, το σύστημα συνιστά εμπορεύματα στα οποία οι πιο κοντινοί γείτονες έδωσαν υψηλή βαθμολογία που δεν έχουν βαθμολογηθεί από τον χρήστη (και με τα οποία προφανώς είναι μη εξοικειωμένος), ένα βασικό ζήτημα είναι πώς μπορεί κανείς να συνδυάσει και να σταθμίσει τις προτιμήσεις του μ' αυτές των γειτόνων του. Ο χρήστης μπορεί να βαθμολογήσει αμέσως τα προτεινόμενα εμπορεύματα, εάν δεν τον ενδιαφέρουν, παρόλα αυτά, συν τω χρόνω το σύστημα αποκτά μια ολοένα και πιο ακριβή αναπαράσταση των προτιμήσεών του.

Καινοτόμα/δημιουργικά συνεργατικά συστήματα Διοίκησης (Collaborative filtering systems) περιλαμβάνονται στα GroupLens (Resnick 1994), στο Bellcore Video Recommender (Hill 1995), και στο Firefly (Shardanand and Maes 1995). Τα συστήματα ποίκιλαν στο πώς στάθμισαν τις εκτιμήσεις των διαφορετικών χρηστών (δηλαδή, στον καθορισμό των γειτόνων και στο κατά πόσο αυτοί είναι κοντινοί) και στον τρόπο με τον οποίο συνδύασαν τις βαθμολογήσεις.

Η Συνεργατική Διοίκηση βρήκε πολλές εφαρμογές στο web. Οι δικτυακοί τόποι ηλεκτρονικού εμπορίου όπως το Amazon.com και το CDNow έχουν σαν χαρακτηριστικό γνώρισμα κέντρα σύστασης, όπου επιπρόσθετα με τις αξιολογήσεις των εμπειρογνομόνων, οι χρήστες μπορούν να αξιολογήσουν εμπορεύματα και στη συνέχεια να λάβουν προσωπικές συστάσεις που υπολογίζονται από έναν μηχανισμό συνεργατικής διοίκησης. Η προτίμηση του χρήστη μπορεί επίσης να εξαχθεί από τη χρήση του site: για παράδειγμα, η αγορά ενός βιβλίου μπορεί να εκληφθεί ως ένδειξη ενδιαφέροντος, όχι μόνο γι' αυτό το βιβλίο, αλλά επίσης και για τον συγγραφέα του βιβλίου.

Η κύρια δύναμη της Συνεργατικής Διοίκησης είναι ότι οι συστάσεις είναι προσωπικές. Στο βαθμό που οι πλησιέστεροι γείτονες κάποιου πραγματικά έχουν παρόμοιο γούστο, μπορεί κανείς να βρεί εμπορεύματα που δεν θα τα είχε σκεφτεί και που είναι αρκετά πιθανό να βρεί ενδιαφέροντα. Δεύτερον, δεν χρειάζεται να ψάχνει κανείς για μια σύσταση ή έναν recommender – απλά αναφέρει τις προτιμήσεις του και λαμβάνει συστάσεις. Τέλος, από υπολογιστική άποψη, η αναπαράσταση των δεδομένων είναι απλή και ομοιόμορφη – ένας πίνακας χρήστη-εμπορευμάτων του οποίου τα κελιά αντιπροσωπεύουν βαθμολογίες – και έτσι είναι επιδεκτικός σε πολλούς διαφορετικούς υπολογιστικούς χειρισμούς.

Η Συνεργατική Διοίκηση δεν υποστηρίζει απλώς μια υπάρχουσα δραστηριότητα. Αντίθετα, απαιτεί οι χρήστες να συμμετάσχουν σε μια νέα υπολογιστικά διαμεσολαβημένη δραστηριότητα. Αυτή η

δραστηριότητα έχει έναν απλό συνδυαστικό ρόλο: αυτόν που ζητά τη σύσταση / αυτόν που παρέχει την προτίμηση-σύσταση (recommendation seeker/preference provider). Αυτό περιγράφεται σαν ομοιομορφία ρόλου. Όλοι κάνουν το ίδιο (βαθμολογούν εμπορεύματα) και απολαμβάνουν τα ίδια οφέλη (λαμβάνουν βαθμολογημένα εμπορεύματα ως συστάσεις), για να πάρει κανείς συστάσεις πρέπει και να δώσει. Αυτό οδηγεί στην αύξηση της γνώσης του συστήματος (και συνεπώς σε καλύτερες συστάσεις), αφού η χρήση της βάσης δεδομένων οδηγεί σε μια ενημερωμένη βάση δεδομένων (Hill et al 1996).

Η ομοιομορφία ρόλου έχει τόσο καλές όσο και κακές πλευρές. Από τη μία πλευρά, η παρατηρηθείσα πρακτική υποδηλώνει ότι οι περισσότεροι άνθρωποι δεν θέλουν να προσφέρουν συστάσεις, αντιθέτως, θέλουν μόνο να τις χρησιμοποιούν. Από την άλλη πλευρά, η βαθμολόγηση εμπορευμάτων δεν είναι τόσο επαχθής δουλειά, και την κάνει κάποιος, ακριβώς όταν θέλει μια σύσταση.

Τέλος, η Συνεργατική Διοίκηση διαχωρίζει την προσωπική επαφή από τη διαδικασία της σύστασης (Hill et al 1996) – δεν χρειάζεται καμία επαφή μεταξύ αυτού που παράγει τη σύσταση και του παραλήπτη της. Φυσικά, αν επιθυμούν οι σχεδιαστές του συστήματος, τα αποτελέσματα του ταιριάσματος μπορούν να χρησιμοποιηθούν για να συστήσουν ανθρώπους στους πιο κοντινούς τους γείτονες. Πράγματι, πρόκειται για μια εξαιρετική τεχνική για τον σχηματισμό κοινότητας, δεδομένου ότι οι άνθρωποι μπορούν να συνδεθούν αυτόματα με άλλους, με τους οποίους μοιράζονται κοινά ενδιαφέροντα.

Υπάρχουν διάφορες τεχνικές προκλήσεις για τους αλγόριθμους συνεργατικής διοίκησης, συμπεριλαμβανομένων των προβλημάτων του "πρώτου βαθμολογητή" και των ελλειπών δεδομένων (sparsity) (Balabanovic, M. και Shoham, Y., Good, N., et al 1999). Δεν μπορεί να προσφερθεί καμία σύσταση για ένα προϊόν μέχρι κάποιος να το βαθμολογήσει. Επιπλέον, στην περίπτωση που ο αριθμός των ατόμων που έχουν βαθμολογήσει τα προϊόντα είναι σχετικά μικρός σε σύγκριση με τον αριθμό των στοιχείων στη βάση δεδομένων, είναι πιθανό ότι δεν θα υπάρχει σημαντική ομοιότητα μεταξύ των χρηστών. Αυτό σημαίνει ότι με τη σειρά τους οι πλησιέστεροι γείτονες πραγματικά δεν θα τόσο κοντά, και έτσι οι συστάσεις δεν θα είναι τόσο καλές. Αυτά τα προβλήματα γίνονται πιο πιεστικά όσο ο αριθμός των εμπορευμάτων αυξάνει.

Μία σημαντική τακτική για την αντιμετώπιση των προβλημάτων αυτών είναι να συνδυάσουμε συστήματα συνεργατικής διοίκησης με συστήματα σύστασης με βάση το περιεχόμενο. Ένα απλό παράδειγμα μπορεί να δείξει τα οφέλη των υβριδικών συστημάτων. Για παράδειγμα, ας υποθέσουμε ότι ένας χρήστης έχει βαθμολογήσει θετικά τη σελίδα του NBA από το ESPN.com, ενώ ένας άλλος έχει βαθμολογήσει θετικά τη σελίδα του NBA από το CNN.com. Το αμιγές σύστημα συνεργατικής διοίκησης δεν θα έβρισκε καμία αντιστοιχία μεταξύ των δύο χρηστών. Ωστόσο, η ανάλυση περιεχομένου μπορεί να αποδείξει ότι τα δύο στοιχεία στην πραγματικότητα είναι πολύ παρόμοια, έτσι, που να δηλώνει ένα ταίριασμα μεταξύ των χρηστών. Το σύστημα Fab (Balabanovic, M. και Shoham, Y.) βασίζεται σε αυτή την αντίληψη. Αυτό, αναλύει το περιεχόμενο των στοιχείων που βαθμολογούν οι χρήστες θετικά για τη δημιουργία προφίλ με βάση το περιεχόμενο των ενδιαφερόντων του χρήστη. Στη συνέχεια εφαρμόζει συνεργατικές τεχνικές φιλτραρίσματος για την αναγνώριση άλλων χρηστών με παρόμοια ενδιαφέροντα.

Άλλοι ερευνητές έχουν αναλύσει θεωρητικά το πρόβλημα των κινήτρων (μια γενίκευση του προβλήματος του πρώτου βαθμολογητή). Και πάλι, το θέμα είναι για ποιο λόγο κάποιος να βαθμολογήσει πρώτος και να μην λάβει κανένα όφελος, όταν μπορεί να περιμένει τους άλλους να αξιολογήσουν κι έτσι να ωφεληθεί από τους άλλους.

Οι Billsus και Pazzani (Billsus, D. & Pazzani, M. 1998) έχουν μια άλλη προσέγγιση για την αντιμετώπιση των προβλημάτων συνεργατικής διοίκησης. Παρατήρησαν ότι το έργο της πρόβλεψης πραγμάτων που μπορεί να αρέσουν σ' ένα χρήστη θα μπορούσε να υλοποιηθεί με ταξινόμηση, ένα πολυερευνημένο θέμα στην κοινότητα της μηχανικής μάθησης. Εφάρμοσαν singular value decomposition στον πίνακα αρχικών βαθμολογιών για την εξαγωγή χαρακτηριστικών και στη συνέχεια, εφάρμοσαν έναν αλγόριθμο μάθησης όπως ένα Νευρωνικό δίκτυο. Αξιοποιώντας τη "λανθάνουσα δομή" στις βαθμολογίες του χρήστη (όπως η λανθάνουσα σημασιολογική ανάλυση (Deerwester, D., et al 1990) εκμεταλλεύεται τη λανθάνουσα δομή στο κείμενο), το σύστημα σε μεγάλο βαθμό μειώνει την ανάγκη για τους χρήστες να

βαθμολογήσουν κοινά στοιχεία πριν να μπορεί ένας χρήστης να χρησιμεύσει σαν προβλέπων για τον άλλο. Πειράματα έδειξαν ότι αυτή η προσέγγιση ξεπέρασε σημαντικά προηγούμενους αλγόριθμους Συνεργατικής διοίκησης.

Πρόσφατα οι Aggarwal et al (Aggarwal, C.A., et al 1999) επινόησαν μια νέα προσέγγιση γράφου της Συνεργατικής Διοίκησης που φαίνεται να αποφεύγει μερικούς από τους περιορισμούς των προηγούμενων αλγορίθμων. Συγκεκριμένα, μπορεί να υπολογίσει ακριβέστερες συστάσεις με λιγότερα δεδομένα.

Όπως κάθε σύστημα που προσφέρει αποτελέσματα στους ανθρώπους βασισμένο σε σημαντική υπολογιστική επεξεργασία, το σύστημα συνεργατικής διοίκησης αντιμετωπίζει το θέμα της εξήγησης – γιατί το σύστημα πιστεύει ότι σε κάποιον θα άρεσε αυτό το στοιχείο;

Ένα τελευταίο σημαντικό θέμα αφορά την έννοια της τύχης στις τυχαίες ανακαλύψεις (serendipity). Δηλαδή, θα θέλαμε ένα σύστημα σύστασης που να "μου λέει κάτι που δεν γνωρίζω ήδη". Πολλά τρέχοντα συστήματα αποτυγχάνουν στη δοκιμή αυτή. Για παράδειγμα, έστω κάποιος που χρησιμοποιεί το κέντρο συστάσεων της Amazon.com. Αφού βαθμολογήσει έναν αριθμό στοιχείων, το σύστημα του συστήνει τον Μάκβεθ του Σαίξπηρ, το οποίο έχει βαθμολογήσει θετικά (με την ένδειξη "είμαι κάτοχος"). Σε αυτό το σημείο, το σύστημα αρχίζει να συστήνει περισσότερα έργα του Σαίξπηρ, όπως Βασιλιάς Λήρ, Άμλετ και Δωδεκάτη Νύχτα. Είναι όμως απίθανο κάποιος που γνωρίζει οποιοδήποτε έργο του Σαίξπηρ να αγνοεί το υπόλοιπο των έργων του. Έτσι, αυτές οι συστάσεις δεν δίνουν νέες πληροφορίες. Τέτοιες καταστάσεις είναι συνηθισμένες. Για να γενικεύσουμε το παραπάνω παράδειγμα με τον Σαίξπηρ, σπάνια είναι χρήσιμο να συστήσεις βιβλία ενός συγγραφέα σε κάποιον, που ήδη έχει βαθμολογήσει βιβλία αυτού του συγγραφέα πολύ θετικά. Το ίδιο μπορεί να γίνει και για τα CD και τους καλλιτέχνες.

Έτσι, ένα σύστημα μπορεί να βελτιωθεί μέσω της γνώσης των συσχετίσεων της γνώμης του χρήστη σχετικά με τα στοιχεία, δηλαδή, αν ένας χρήστης έχει γνωμοδοτήσει σχετικά με το στοιχείο X, είναι πολύ πιθανό ότι ο ίδιος έχει ήδη γνωμοδοτήσει σχετικά με το στοιχείο Y. Μία προσέγγιση σε αυτό το πρόβλημα είναι η δημιουργία γνώσης σχετικά με τα στοιχεία, δημιουργώντας ένα υβριδικό σύστημα με βάση τόσο το περιεχόμενο, όσο και τη συνεργατική διοίκηση.

2.3 Διαμόρφωση και στήριξη κοινοτήτων ενδιαφέροντος

Μια πρώτη πρόκληση είναι να μελετήσουμε, να διδαχθούμε και να υποστηρίξουμε με τον καλύτερο τρόπο τις κοινότητες που σχηματίζονται φυσικά. Για παράδειγμα, όπου έχουμε μιλήσει για κάποιους ρόλους – ο αιτών τη σύσταση (recommendation seeker), ο παρέχων τη σύσταση (recommender) και ο παρέχων την προτίμηση (preference provider) – και τους αντιμετωπίσαμε ως διακριτούς, καμιά από αυτές τις υποθέσεις δεν ισχύει στις πραγματικές κοινότητες. Οι παρατηρήσεις των ιστολογίων το καταδεικνύουν.

Το Slashdot.com είναι ένα πολύ γνωστό μπλογκ / μια web κοινότητα του οποίου το σύνθημα είναι "νέα για σπασίκλες". Θέματα όπως το Linux, η Java και το λογισμικό ανοικτού κώδικα είναι τα βασικά ενδιαφέροντα. Το Slashdot ξεκίνησε από λίγα άτομα ως τόπος για τη συλλογή και συζήτηση πληροφοριών που εύρισκαν ενδιαφέρουσες. Μεγάλωσε ραγδαία και σύντομα έπεσε θύμα των πολλών προβλημάτων που προκαλεί η υπερφόρτωση των πληροφοριών που ένα μπλογκ προσπαθεί να αποφύγει. Δεκάδες ιστορίες και χιλιάδες σχόλια καταχωρούνται ημερησίως – πάρα πολλά για να τα διαβάσει κανείς.

Ο εκδότης του Slashdot ανέπτυξε έναν ενδιαφέροντα μηχανισμό (<http://Slashdot.com/moderation.shtml>) για την αντιμετώπιση του προβλήματος. Οι «καλοί» συμμετέχοντες στον τόπο έχουν περιορισμένες αρμοδιότητες διαχείρισης. Για περιορισμένο χρόνο μπορούν να βαθμολογούν μερικά σχόλια ως καλά ή κακά, αυξάνοντας έτσι ή ελαττώνοντας μια βαθμολογία. Οι αναγνώστες μπορούν να ορίσουν φίλτρα για να δουν μόνο περιεχόμενο με συγκεκριμένη βαθμολογία. Το Slashdot είναι μια κοινότητα με πολλούς, μεταβατικούς και μεταβαλλόμενους ρόλους. Αντί για ένα σύνολο συστάσεων που παράγουν λίγα άτομα και τα οποία καταναλώνονται από άλλους, χρησιμεύει σαν το σημειωματάριο μιας κοινότητας, ένα μέσο στο οποίο όλοι οι συμμετέχοντες μπορούν να προτείνουν και να σχολιάσουν ιδέες.

Οι συμμετέχοντες στις κοινότητες αυτές μπορεί να χρειαστούν συστήματα σύστασης. Εάν κάποιος θέλει να διατηρήσει μια λίστα FAQ ή εάν συνεισφέρει σε ένα μπλογκ, είναι αδύνατον να διαβάσει και να αξιολογήσει όλες τις πληροφορίες του web για ένα οποιοδήποτε αρκετά ευρύ θέμα. Ένας άνθρωπος που κάνει συστάσεις (recommender), χρειάζεται ένα σύστημα σύστασης.

Διαφορετικοί τύποι συστημάτων σύστασης θα μπορούσαν να χρησιμοποιηθούν για να προτείνουν περιεχόμενο. Ένα content-based σύστημα σύστασης θα μπορούσε να καταγράφει τα γραπτά που ένας εκδότης διαβάζει, προκειμένου να τα εκδώσει, να σημειώνει αυτά που επιλέγει ή απορρίπτει, και σταδιακά να εξελίξει ένα φίλτρο για να καταγράψει αυτές τις προτιμήσεις. Ένα σύστημα εξόρυξης κοινωνικών δεδομένων μπορεί να επεξεργάζεται σχετικές ομάδες συζήτησης ή συνεχώς να ανιχνεύει και να αναλύει τις τοποθεσίες web για νέα ή για δημοφιλή προϊόντα. Με ένα κοινωνικό σύστημα σύστασης, ο εκδότης αποκτά πρόσβαση στις απόψεις πολλών διαφορετικών ατόμων, αυτό είναι ένα είδος "παλμού της κοινότητας". Έτσι, μπορεί να έρθει σε επαφή με νέες πληροφορίες και ιδέες. Από την άλλη, με ένα σύστημα σύστασης με βάση το περιεχόμενο ο εκδότης θα πάρει προτάσεις για στοιχεία που είναι παρόμοια με αυτά που έχει επιλέξει στο παρελθόν. Αυτό μπορεί να οδηγήσει σε μια πιο συνεκτική, αλλά στενή και στατική προσφορά των πληροφοριών.

2.3.1 Διαμόρφωση Κοινοτήτων με προστασία του ιδιωτικού απορρήτου

Τα Συστήματα σύστασης μπορούν να συνδέσουν άτομα με βάση κοινά ενδιαφέροντα. Τα συστήματα που εξορύσσουν προτιμήσεις από εγγραφές δραστηριοτήτων μπορούν να επιλέξουν να δημοσιοποιήσουν την ταυτότητα των ατόμων που παράγουν τις προτιμήσεις. Τα συστήματα συνεργατικής διοίκησης μπορεί να επικοινωνήσουν το σύνολο των "γειτόνων" που χρησιμοποιήθηκαν για τον υπολογισμό μιας σύστασης. Σε οποιαδήποτε περίπτωση, οι χρήστες του συστήματος έχουν τη δυνατότητα να επικοινωνήσουν και να

σχηματίσουν μια κοινότητα με άλλους χρήστες που έχουν κοινά ενδιαφέροντα. Ωστόσο, αυτή η δυνατότητα θέτει σημαντικές ανησυχίες σχετικά με το απόρρητο.

Οι ανησυχίες αυτές είναι πιο έντονες για συστήματα εξόρυξης κοινωνικών δεδομένων, επειδή εξάγουν πληροφορίες από το αρχικό περιβάλλον. Το PHOAKS για παράδειγμα: εξάγει προτιμήσεις σχετικά με ιστοσελίδες από μηνύματα ομάδων συζήτησης, και συγκεντρώνει αυτές τις προτιμήσεις σε συστάσεις που διατίθενται σε μια τοποθεσία web. Προφανώς, οι περισσότεροι χρήστες που επισκέπτονται την τοποθεσία web δεν ήταν οι συμμετέχοντες στην αρχική ομάδα συζήτησης. Ούτε ο τρόπος με τον οποίο έγινε η επεξεργασία των πληροφοριών, ούτε το νέο κοινό που μπορεί να έχει πρόσβαση στα αποτελέσματα ήταν εν γνώσει ή είχε προβλεφθεί από τους παραγωγούς των πληροφοριών.

Ένας σχεδιαστής συστήματος έχει διάφορες επιλογές για την εξισορρόπηση του απορρήτου των μεμονωμένων ατόμων και των ομάδων ενάντια στις ευκαιρίες για εκτεταμένες επαφές μεταξύ των ανθρώπων. Πρώτον, μπορεί να παρουσιάζονται μόνο οι πληροφορίες που έχουν συγκεντρωθεί ανεξάρτητα από το πλαίσιο. Για το PHOAKS, αυτό θα μπορούσε να σημαίνει να παρουσιάζει μόνο καταλόγους των προτεινόμενων διευθύνσεων URL, χωρίς πληροφορίες σχετικά με τα μηνύματα σύστασης ή τα πρόσωπα που καταχώρησαν τις συστάσεις. Ωστόσο, αυτό οδηγεί σε λιγότερο πλούσιες και ενημερωτικές συστάσεις και δεν δίνει καμιά ευκαιρία για τους ανθρώπους να κάνουν νέες γνωριμίες με άλλα άτομα με τα οποία μοιράζονται κοινά ενδιαφέροντα. Δεύτερον, κάποιος μπορεί να τα κάνει όλα ρητά /άμεσα. Για παράδειγμα, το PHOAKS θα μπορούσε να έχει σχεδιαστεί για να κάνει εμφανή την ταυτότητα των recommenders και να καθιστά τεχνικά εύκολη την επικοινωνία με έναν recommender (συμπεριλαμβανοντας συνδέσεις mailto:). Αντ'αυτού συμπεριλήφθηκαν οι διευθύνσεις e-mail των recommenders, αλλά για να τις βρει κάποιος θα πρέπει να ψάξει λίγο, και στη συνέχεια μπορεί να στείλει mail.

Γενικότερα, ενδιαφέρουν μέσες λύσεις που βρίσκονται μεταξύ των δύο άκρων: της πλήρους γνωστοποίησης και της πλήρους ανωνυμίας. Ένα καλό σημείο για να ξεκινήσει κανείς είναι με τεχνικές που χρησιμοποιούνται σε τοποθεσίες που παρέχουν υπηρεσίες γνωριμιών. Στις περιπτώσεις αυτές, το σύστημα είναι ένας αξιόπιστος ενδιάμεσος φορέας, που διαμεσολαβεί την αλληλεπίδραση μεταξύ των ανθρώπων. Οι συμμετέχοντες μπορούν σταδιακά να αποκαλύπτουν περισσότερα σχετικά με τους εαυτούς τους, αρχίζοντας με το login τους, μετά με την προσωπική τους διεύθυνση ηλεκτρονικού ταχυδρομίου και στη συνέχεια με άλλες πληροφορίες καθώς νοιώθουν περισσότερο άνετα.

2.3.2 Συνδυάζοντας πολλούς τύπους πληροφοριών για τον υπολογισμό συστάσεων

Η βάση των αλγορίθμων συνεργατικής διοίκησης είναι το ταίριασμα ατόμων με βάση παρόμοια ενδιαφέροντα. Η λήψη συστάσεων από κάποιον με παρόμοιες προτιμήσεις είναι ένα καλό ξεκίνημα, μπορεί όμως κάποιος να θέλει κάτι άλλο: το πρόσωπο που πραγματοποιεί τη σύσταση να είναι ένας εμπειρογνώμων σχετικά με ένα θέμα. Ίσως, κάποιος να προτιμάει να πάρει μια σύσταση με βάση τη γνωμοδότηση ενός εμπειρογνώμονα, αντί για αυτή 10 ατόμων, ακόμη και όταν αυτοί οι 10 έχουν πράγματι ενδιαφέροντα κάπως πιο κοντά στα δικά του.

Αυτό δημιουργεί πολλές προκλήσεις, συμπεριλαμβανομένης της απόκτησης πληροφορίας εμπειρογνωμοσύνης, ελέγχου της καταλληλότητας της πληροφορίας και του συνδυασμού της με πληροφορίες σχετικά με την ομοιότητα των προτιμήσεων για τον υπολογισμό της σύστασης. Διάφορες τεχνικές για την λήψη πληροφοριών σχετικά με την εμπειρία μπορούν να διερευνηθούν. Για παράδειγμα, σε μια ηλεκτρονική συνομιλία, μετρήσεις όπως το ποσό των μηνυμάτων με τα οποία ένα πρόσωπο συμβάλλει, καθώς και ο αριθμός των απαντήσεων που λαμβάνει θα μπορούσαν να χρησιμοποιηθούν για την εκτίμηση της εμπειρίας ενός συμμετέχοντος.

Σε ακαδημαϊκό πλαίσιο, θα μπορούσαν να χρησιμοποιηθούν οι μέθοδοι ευρετηρίασης των παραπομπών (citation indexing). Μια περαιτέρω σκέψη είναι ότι η πείρα είναι συγκεκριμένη πάνω σε ένα θέμα – για παράδειγμα, στον τομέα της μουσικής, ένα άτομο μπορεί να είναι ειδήμων στην Μπαρόκ Συστήματα Σύστασης βασισμένα σε Τεχνικές Νοημοσύνης Σμήνους

μουσική, άλλο στη jazz, και ένα τρίτο στη rock. Διάφορες τεχνικές μπορούν να βοηθήσουν την κατηγοριοποίηση της εμπειρίας του ατόμου. Εάν ένα σύστημα έχει κατηγορηματική πληροφορία σχετικά με στοιχεία (π.χ., το είδος τους), τότε αν ένα πρόσωπο βαθμολογεί συχνότερα στοιχεία από μία κατηγορία (και πιο πολύ) από ότι άλλα είδη, αυτό μπορεί να υποδεικνύει εμπειρογνωμοσύνη σε αυτήν την κατηγορία. Σε μια εφαρμογή συνομιλίας, τα μηνύματα που παράγει ένα άτομο μπορούν να αναλυθούν για την εξαγωγή θεμάτων που συζητά συχνά. (Κατανοούμε ότι το να συζητά κάποιος ένα θέμα πολύ, δεν τον κάνει απαραίτητα έναν εμπειρογνώμονα, αλλά αυτό είναι ένα καλό σημείο για να ξεκινήσει ο πειραματισμός.) Το τελευταίο ζήτημα είναι πώς να συνδυαστούν οι πληροφορίες εμπειρογνωμοσύνης και προτίμησης. Οι υπάρχοντες αλγόριθμοι συνεργατικής διοίκησης θα μπορούσαν απλά να διαμορφωθούν με έναν άλλο όρο που σταθμίζει την επιρροή των γειτόνων σε μια σύσταση βάσει των γνώσεών τους. Ωστόσο, τα βάρη δεν είναι σαφή εκ των προτέρων, είναι αναγκαία πειράματα, και ίσως διαφορετικοί συνδυασμοί θα ήταν κατάλληλοι υπό διαφορετικές συνθήκες.

2.3.3 Συνδυάζοντας πολλαπλές πηγές προτιμήσεων

Η σκέψη για τον συνδυασμό των πληροφοριών εμπειρογνωμοσύνης και προτίμησης οδηγεί σε μια άλλη υλοποίηση – οι προτιμήσεις μπορούν να αποκτηθούν από διάφορες πηγές, και οι αλγόριθμοι πρέπει να είναι σε θέση να συνδυάσουν διαφορετικούς τύπους προτιμήσεων κατάλληλα. Για παράδειγμα, οι προτιμήσεις στο PHOAKS έχουν ληφθεί από την εξόρυξη και συγκέντρωση γνωμοδοτήσεων από τα μηνύματα του Usenet. Ωστόσο, οι χρήστες του PHOAKS ήταν σε θέση να αξιολογήσουν τις συνιστώμενες διευθύνσεις URL, και το σύστημα κατέγραφε το ιστορικό χρήσης (δηλαδή, τη συχνότητα επισκέψεων της κάθε διεύθυνσης URL). Έτσι, σε τελική ανάλυση, υπήρχαν τρεις πηγές προτιμήσεων σχετικά με τα URL: αναφορές σε μηνύματα ομάδων συζήτησης, το ιστορικό χρήσης και οι ρητές/άμεσες αξιολογήσεις.

Πώς μπορούν να συνδυαστούν οι τρεις τύποι των προτιμήσεων είναι μια πρόκληση: δεν είναι σαφές πόσο βάρος να αντιστοιχηθεί σε έναν συγκεκριμένο τύπο. Η ανάλυση του ιστορικού χρήσης, ιδίως, απαιτεί κάποια σκέψη. Για παράδειγμα, απλά η καταμέτρηση του πόσες φορές οι χρήστες έχουν κάνει κλικ σε κάθε διεύθυνση URL σαν ένα μέτρο προτίμησης είναι μια προφανής στρατηγική. Ωστόσο, θα περίμενε κανείς οι χρήστες να κάνουν κλικ περισσότερο συχνά στις διευθύνσεις URL που ήταν ψηλότερα στη λίστα εμφάνισης. Με βάση αυτή τη γνώση, η απόκλιση από το αναμενόμενο μοτίβο – δηλαδή, οι διευθύνσεις URL που έχουν σχετικά περισσότερα ή λιγότερα κλικ από τα αναμενόμενα – μάλλον θα πρέπει να εξετάζεται.

Αυτό το παράδειγμα παρουσιάζει επίσης μια ενδιαφέρουσα γενική παρατήρηση: η χρήση μιας σύστασης για τη λήψη μιας απόφασης επίσης μπορεί να δώσει πρόσθετα δεδομένα προτίμησης, που μπορούν να χρησιμοποιηθούν για να εξελιχθεί η σύσταση. Και ένας χρήστης μπορεί να παίξει πολλούς ρόλους ταυτόχρονα – αυτού που ζητά τη σύσταση και του δημιουργού της προτίμησης – .

2.4 Συμπεράσματα

Τα Συστήματα σύστασης αναπτύχθηκαν σε απάντηση μιας προφανούς ανάγκης: να βοηθηθούν οι άνθρωποι να αντιμετωπίσουν τον κόσμο της αφθονίας και της υπερφόρτωσης των πληροφοριών. Επιπλέον, κατέστη σαφές ότι μπορούν να συνδέουν άτομα με άλλα άτομα που μοιράζονται κοινά ενδιαφέροντα. Υπάρχει ένα σύνολο από τέσσερα σημαντικά ζητήματα για τα Συστήματα σύστασης:

(1) ο τρόπος με τον οποίο λαμβάνονται και χρησιμοποιούνται τα δεδομένα προτίμησης, (2) ο ρόλος που παίζουν οι άνθρωποι και οι υπολογισμοί, καθώς και οι σχετικοί τύποι επικοινωνίας, (3) οι αλγόριθμοι για τη σύνδεση των ατόμων και τον υπολογισμό των συστάσεων και (4) η παρουσίαση των συστάσεων στους χρήστες. Στη συνέχεια, προσδιορίστηκαν τέσσερις μεγάλες προσεγγίσεις για τα Συστήματα σύστασης, που μπορεί να διακρίνονται σε μεγάλο βαθμό από τα ζητήματα που αντιμετωπίζουν και τον τρόπο με τον οποίο τα αντιμετωπίζουν.

ΚΕΦΑΛΑΙΟ 3

3. Συστήματα Σύστασης βασισμένα σε Τεχνικές Νοημοσύνης Σμήνους

3.1 Εισαγωγή

Η συνεχόμενη ανάπτυξη και η αύξηση της πολυπλοκότητας των εφαρμογών που βασίζονται στο Web, από το παραδοσιακό e-commerce στις υπηρεσίες Web, σε όλα τα είδη δυναμικού περιεχομένου, οδήγησε σε πολλαπλασιασμό των εργαλείων αναζήτησης. Προσωπικές υπηρεσίες, όπως τα συστήματα σύστασης, βοηθούν στην προσέλκυση επισκεπτών, στη μετατροπή περιστασιακών περιηγητών ιστοσελίδων σε πελάτες, ή βοηθούν τους επισκέπτες να εντοπίσουν αποτελεσματικότερα σχετικές πληροφορίες. Ο στόχος της κάθε σύστασης σε οποιοδήποτε πεδίο είναι να κάνει μια επιλογή από όλα τα πιθανά στοιχεία με τη χρήση ορισμένων χαρακτηριστικών προκαθορισμένων από το περιβάλλον. Οι προτιμήσεις των πελατών προς συγκεκριμένα χαρακτηριστικά των προϊόντων (από βιβλία έως αντικείμενα μάθησης) αναλύονται με διαφορετικές τεχνικές και στη συνέχεια προκύπτουν κανόνες σχετικά με το προφίλ των ενδιαφερόντων του πελάτη μέσα σε ένα συγκεκριμένο πλαίσιο, προκειμένου να γίνει η σύσταση (Berlanga et al, 2005). Έτσι, σε γενικές γραμμές, οι διαδικασίες σύστασης ορίζονται ως το αποτέλεσμα πολλών μικρών παραμέτρων όλων μαζί σε συνεχή αλληλεπίδραση.

Σημαντικά χαρακτηριστικά πολύπλοκων συστημάτων έχουν εντοπιστεί στα Συστήματα Σύστασης. Η συμπεριφορά του χρήστη στη διαδικασία λήψης αποφάσεων δεν υπολογίζεται με συνολικό έλεγχο, αλλά αντίθετα από αλληλεπιδράσεις με το περιβάλλον. Ο χρήστης έχει να αντιμετωπίσει ένα εκτεταμένο ποσό αποσπασματικών πληροφοριών. Και τα δύο συστήματα επομένως, είναι καταναμημένα, μεγάλα, ανοικτά και ετερογενή.

Η διαδικασία σύστασης στο ηλεκτρονικό εμπόριο είναι το καθήκον της επιλογής και της οργάνωσης των υπηρεσιών σύμφωνα με το τί ενδιαφέρει τον χρήστη. Η προσέγγιση των Gil και García-Reñano χρησιμοποιεί οντολογική γνώση πεδίου για να εκτελέσει αυτή την εργασία. Οι οντολογίες πεδίου είναι πολύτιμα εργαλεία στην παρουσίαση υπηρεσιών ρητής γνώσεως και στην εξαγωγή σχετικής σημασιολογίας πεδίου στο προφίλ του χρήστη. Οι υπηρεσίες είναι σημασιολογικά ορισμένες, κάτι το οποίο παρέχει όλες τις πληροφορίες που χρειάζονται οι υπηρεσίες Web για να αλληλεπιδρούν και έτσι, μπορούν να μετατραπούν σε πράκτορες με σημασιολογικά χαρακτηριστικά.

Έτσι, μια πολύπλοκη δυναμική διαδικασία στη σύσταση μετατρέπεται σε απλές αλληλεπιδράσεις πρακτόρων στο περιβάλλον. Παρακάτω θα παρουσιαστεί μια αρχιτεκτονική που βασίζεται σε σχέσεις μεταξύ recommenders και σύνθετα συστήματα για την υποστήριξη της σύστασης σε e-learning σύστημα που δουλεύει με αντικείμενα μάθησης (Learning Objects - LOs).

Πριν από την εμφάνιση της Service Oriented Architecture όλη η πληροφορία βρισκόταν σε στατικές σελίδες στο Internet. Έτσι, οι μηχανές αναζήτησης συχνά έβρισκαν χιλιάδες δυνητικά σχετικές τοποθεσίες. Για ορισμένες εφαρμογές ένας χρήστης ήταν υποχρεωμένος να ορίσει τους στόχους του σε όρους ενός ερωτήματος, το οποίο συγκρίνονταν στη συνέχεια (συνήθως σε επίπεδο μιας απλής λέξης-κλειδί) με έγγραφα σε μια συλλογή και συγκεκριμένα με εκείνα που ενδέχονταν να είναι πιο σχετικά με το ερώτημα και ενδεχομένως να αφορούσαν τον χρήστη. Στην κοινότητα της Τεχνητής Νοημοσύνης (AI), έχει γίνει πολλή δουλειά σχετικά με τον τρόπο με τον οποίο μπορεί να βοηθήσει η AI για την επίλυση αυτού του προβλήματος. Οι έννοιες προσωπικών μηχανών αναζήτησης, πράκτορες ευφυούς λογισμικού, και συστήματα σύστασης έχουν αποκτήσει μεγάλη αποδοχή μεταξύ των χρηστών για την βοήθειά τους στην αναζήτηση, διαλογή, ταξινόμηση, φιλτράρισμα και στην κοινή χρήση του τεράστιου ποσού πληροφορίας διαθέσιμου στο Web. Ο συνδυασμός της μοντελοποίησης των προτιμήσεων των

συγκεκριμένων χρηστών, η δημιουργία μοντέλων περιεχομένου και η μοντελοποίηση κοινωνικών μοτίβων σε ευφυείς πράκτορες παρέχει στους χρήστες τα μέσα για τη διαχείριση πληροφοριών με ένα ορθολογικό τρόπο και έτσι τους βοηθά να ξεπεράσουν την υπερφόρτωση πληροφοριών.

Σήμερα, χάρη στην ανάπτυξη του Σημασιολογικού Ιστού (Semantic Web), υπάρχει η απαίτηση για πρόσβαση όχι μόνο σε στατικά έγγραφα, τα οποία συγκεντρώνουν χρήσιμες πληροφορίες, αλλά και σε υπηρεσίες που παρέχουν νέους τρόπους για να προσφέρουν πληροφορίες, και για τις οποίες πρέπει να παρασχεθεί ένα νέο μοντέλο διαδικασιών. Έτσι, προέκυψε η τεχνολογία ανάκτησης υπηρεσιών, αλλά η κοινότητα ανάκτησης πληροφορίας έχει επικεντρωθεί στην ανάκτηση των εγγράφων, και έχει ως αποτέλεσμα μια προσέγγιση με βάση μια λέξη-κλειδί. Καθώς το πλήθος των εν λόγω υπηρεσιών αυξάνεται, γίνεται όλο και πιο σημαντικό να παρέχουν εργαλεία που επιτρέπουν στα άτομα (και στο λογισμικό) να βρίσκουν γρήγορα τις υπηρεσίες που χρειάζονται με προσοχή στην εξατομικευμένη επιλογή.

3.2 Recommenders στο e-Commerce

Αυτά τα εργαλεία στο περιβάλλον του e-Commerce (EC) λειτουργούν σαν ένας εξειδικευμένος πωλητής για κάθε πελάτη, και συνδέονται συνήθως με τις δυνατότητες εξατομίκευσης για κάθε χρήστη, με βάση την ανάλυση των προτιμήσεων και των ενδιαφερόντων του. Οι recommenders βασίζονται κυρίως στις διεπαφές χρήστη, σε τεχνικές του μάρκετινγκ και σε μεγάλες ποσότητες πληροφορίας σχετικά με άλλους πελάτες και προϊόντα, προκειμένου να προσφέρουν το σωστό είδος στον σωστό πελάτη. Οι recommenders είναι τα βασικά στοιχεία στην εξασφάλιση της χρηστικότητας και της εμπιστοσύνης σε έναν δικτυακό τόπο, γεγονός που σημαίνει ότι παίζουν ένα σημαντικό ρόλο στη σχεδίαση οποιασδήποτε αγοράς. Οι EC recommenders γίνονται σταδιακά ισχυρά εργαλεία για το ηλεκτρονικό εμπόριο καλύπτοντας πολύπλοκους μηχανισμούς κυρίως για την υποστήριξη της διαδικασίας λήψης αποφάσεων ενός χρήστη, επιτρέποντας στα ανθρώπινα όντα να χρησιμοποιήσουν αναλογική συλλογιστική. Έτσι, οι recommenders στο EC πρέπει να αναπτυχθούν από κοινού με πολλούς τομείς, όπως είναι η αλληλεπίδραση ανθρώπου - υπολογιστή (Human-Computer Interaction), η εξόρυξη γνώσης από δεδομένα, οι γνωστικές επιστήμες ή το μάρκετινγκ. Οι τοποθεσίες EC έκαναν μια μεγάλη προσπάθεια για να παράσχουν στους πελάτες τους εργαλεία για να κάνουν τις αγορές τους στο Διαδίκτυο ευκολότερες. Η ανάγκη να γίνουν οι τοποθεσίες φιλικές προς το χρήστη προϋποθέτει αναγκαστικά την κατανόηση της συμπεριφοράς των καταναλωτών, προκειμένου να διευκολύνει και να εξατομικεύσει την πρόσβαση στον μεγάλο όγκο των πληροφοριών που πρέπει να αναζητηθεί και να αφομοιωθεί πριν από την κάθε αγορά.

Υπάρχει ένας μεγάλος αριθμός recommenders με πτυχές εξατομίκευσης στο Internet. Μια πλήρης επισκόπηση των recommenders στην EC, υπάρχει στα άρθρα των (Sarwar et al, 2000), (Shafer et al, 2001) και (Montaner et al, 2003). Μια πρόχειρη ταξινόμηση μπορεί να γίνει με βάση το είδος της πληροφορίας και του τρόπου με τον οποίο το σύστημα σύστασης χειρίζεται αυτή την πληροφορία για να λειτουργήσει. Λαμβάνοντας υπόψη το σύστημα αγορών, τις κοινότητες καταναλωτών ή ένα υβρίδιο αυτών των δύο ως το κύριο στοιχείο για τη δημιουργία της σύστασης, διακρίνουμε τρεις κατηγορίες.

1. Τα **Collaborative-social-filtering systems**: Τα συστήματα συνεργατικής διοίκησης οικοδομούν τη σύσταση από το άθροισμα των προτιμήσεων των καταναλωτών. Τέτοιου είδους συστήματα κάνουν ταιριάσματα με άλλους χρήστες βασιζόμενα στην ομοιότητα των συμπεριφορικών ή των κοινωνικών προτύπων.

Η στατιστική ανάλυση στην εξαγωγή δεδομένων ή οι τεχνικές εξόρυξης δεδομένων και ανακάλυψης της γνώσης σε βάσεις δεδομένων (KDD) (παρακολούθηση της συμπεριφοράς του χρήστη εντός του συστήματος, αξιολόγηση των υπηρεσιών, ιστορικό των αγορών, κλπ.) δημιουργούν τη σύσταση αναλογικά με πολλούς άλλους χρήστες. Οι ομοιότητες μεταξύ των χρηστών υπολογίζονται χρησιμοποιώντας τη συσχέτιση χρήστη προς χρήστη. Η τεχνική αυτή βρίσκει ένα σύνολο "πλησιέστερων γειτόνων" για κάθε χρήστη προκειμένου να προσδιοριστούν παρόμοιες προτιμήσεις.

Η τεχνική αυτή πάσχει κυρίως από το πρόβλημα των ελλιπών δεδομένων (sparsity) λόγω της ανάγκης για ένα μεγάλο όγκο χρηστών σε σχέση με τον όγκο των προσφερόμενων στοιχείων (κρίσιμη μάζα) με σκοπό την παροχή κατάλληλων προτάσεων. Είναι επίσης αδύνατη γι' αυτούς η προσφορά νέων υπηρεσιών επειδή δεν υπάρχουν εγγραφές προηγούμενων αγορών τους.

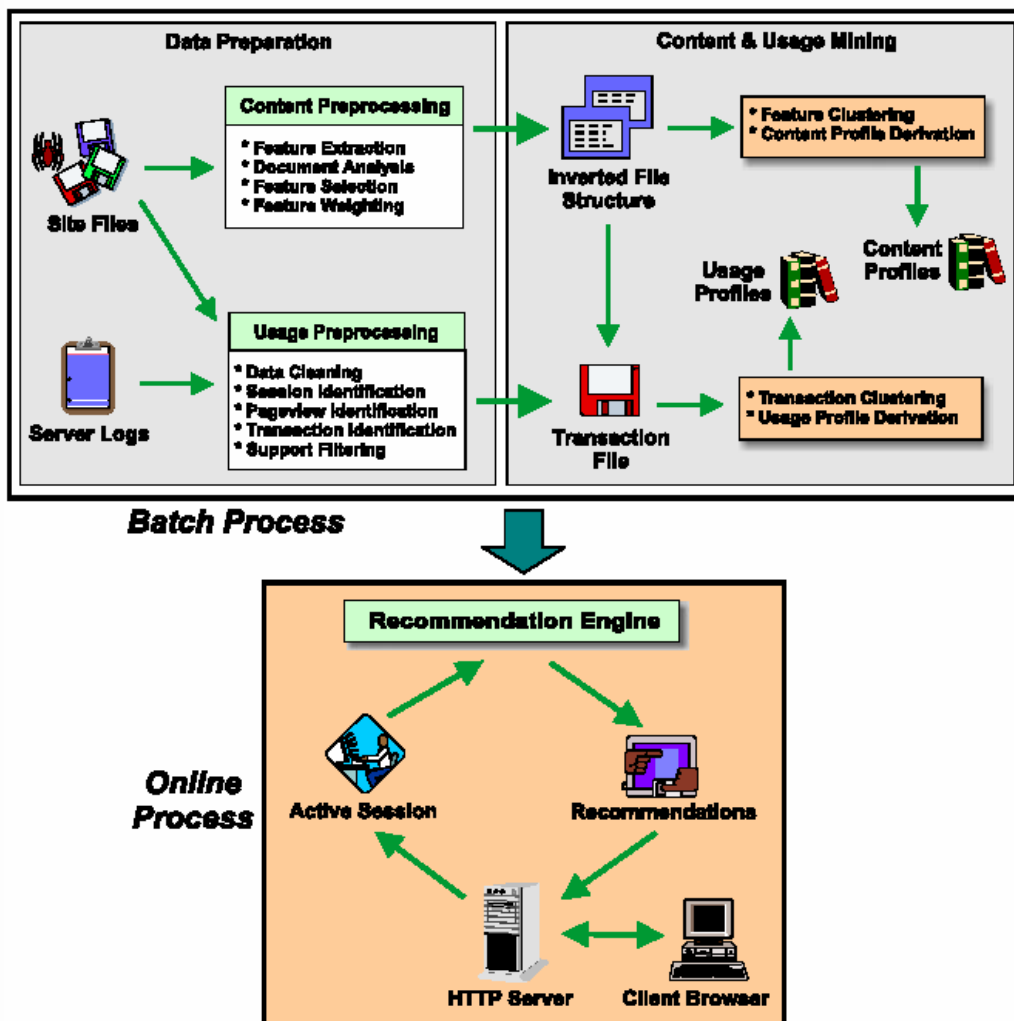
2. **Content-based-filtering systems:** Τα Συστήματα με βάση το περιεχόμενο εξάγουν τις πληροφορίες για τις συστάσεις με βάση τα στοιχεία που έχει αγοράσει στο παρελθόν ο χρήστης. Τέτοιου είδους συστήματα χρησιμοποιούν μηχανική μάθηση με επιτήρηση, για να παρέξει έναν ταξινομητή για τη διάκριση μεταξύ των υπηρεσιών που θα είναι ενδιαφέρουσες ή μη για τον χρήστη λόγω του ιστορικού των αγορών του. Οι ταξινομητές μπορούν να υλοποιηθούν χρησιμοποιώντας πολλές διαφορετικές τεχνικές από την τεχνητή νοημοσύνη ως τα νευρωνικά δίκτυα, τα Bayesian δίκτυα, τους επαγωγικούς κανόνες, τα δένδρα αποφάσεων κ.λπ. Το μοντέλο χρήστη που αντιπροσωπεύεται από τον ταξινομητή που επιτρέπει στο σύστημα να σταθμίζει την προτίμηση ή την απαρésκεια για ένα στοιχείο. Αυτή η πληροφορία προσδιορίζει τα πιο σταθμισμένα στοιχεία που τελικά συνιστώνται στον χρήστη. Ορισμένα συστήματα με βάση το περιεχόμενο χρησιμοποιούν επίσης τη συσχέτιση στοιχείου με στοιχείο για τον προσδιορισμό κανόνων συσχέτισης μεταξύ των στοιχείων. Η πιο πάνω τεχνική πάσχει κυρίως από το πρόβλημα της υπερ-εξειδίκευσης επειδή ο καταναλωτής οδηγείται στο να αγοράσει τα ίδια είδη στοιχείων που έχει ήδη αγοράσει. Αυτό είναι ένα πρόβλημα επίσης για νέα είδη στο κατάστημα (κανένας δεν έχει αγοράσει αυτό το είδος πριν).

3. **Knowledge-based systems:** Τα βασισμένα στη γνώση συστήματα μπορούν να βοηθούν ως ένα υβρίδιο μεταξύ συστημάτων Συνεργατικής διοίκησης και συστημάτων με βάση το περιεχόμενο, τη λειτουργία των οποίων επίσης ενισχύει.

Τα συστήματα αυτά χτίζουν γνώσεις σχετικά με τους χρήστες και συνδέονται επίσης με τη γνώση των υπηρεσιών. Αυτή η πληροφορία χρησιμοποιείται για να αιτιολογήσει γιατί το στοιχείο αυτό ανταποκρίνεται στις απαιτήσεις του χρήστη. Η σχέση μεταξύ των υπηρεσιών και των πελατών οδηγεί στην εξαγωγή συμπερασμάτων που δημιουργούν τη γνώση στον EC μηχανισμό. Ορισμένα από αυτά τα συστήματα παρέχουν νέες λύσεις για το φιλτράρισμα των πληροφοριών με βάση στην εμπιστοσύνη.

3.2.1 Τεχνολογίες της Σύστασης

Καθώς το πρόβλημα της σύστασης είναι πολύ διαδεδομένο, ανταποκρινόμενοι στις διάφορες πτυχές του και προκειμένου να το επιλύσουμε, πρέπει να προσδιορίσουμε δύο κύριους τεχνολογικούς τρόπους για να το προσεγγίσουμε: εξόρυξη του Web και τεχνολογίες που βασίζονται σε πράκτορες. Και οι δύο αυτοί τρόποι χρησιμοποιούνται επίσης στον τομέα της Τεχνητής Νοημοσύνης.

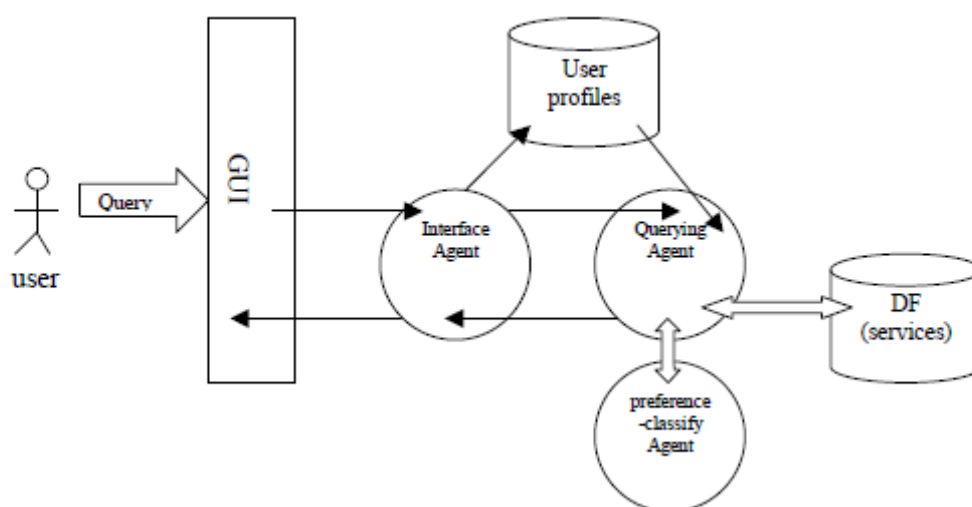


Σχήμα 1: Γενικό πλαίσιο για την αυτόματη εξατομίκευση μέσω εξόρυξης Web (Mobasher et al, 2000)

Ο όρος εξόρυξη του Web (Web mining) είναι η εξόρυξη ενδιαφέρουσας και χρήσιμης γνώσης, καθώς και των έμμεσων πληροφοριών από δραστηριότητες που συνδέονται με το Web. Οι Web διακομιστές καταγράφουν και συλλέγουν δεδομένα σχετικά με τις αλληλεπιδράσεις του χρήστη κάθε φορά που λαμβάνονται αιτήσεις για πόρους.

Η ανάλυση των αρχείων καταγραφής της πρόσβασης στο Web μπορεί να συμβάλει στην κατανόηση της συμπεριφοράς του χρήστη. Τα προφίλ των χρηστών έχουν δημιουργηθεί από συνδυασμό των διαδρομών πλοήγησης του χρήστη με άλλα χαρακτηριστικά των δεδομένων, όπως η διάρκεια επίσκεψης μιας σελίδας, η δομή των υπερσυνδέσμων και το περιεχόμενο της σελίδας.

Ένας άλλος δρόμος στην ανάκτηση πληροφορίας είναι μέσω εφαρμογών βασισμένων σε πράκτορες για το φιλτράρισμα και την παρουσίαση σχετικής πληροφορίας για τον χρήστη, εφαρμογές που βασίζονται στα οικοσυστήματα των προσαρμοστικών συστημάτων πολλών πρακτόρων (adaptive multiagent systems). Στο παρακάτω σχήμα παρουσιάζεται μια κλασική προβολή της αρχιτεκτονικής του συστήματος πρακτόρων.



Σχήμα 2: Αρχιτεκτονική του συστήματος recommender βασισμένη σε πράκτορες

Οι Sheth και Maes, το 1993 εφάρμοσαν μια αρχιτεκτονική οικοσυστήματος πρακτόρων για να φιλτράρουν τις ειδήσεις στο Internet σε ένα σύστημα που ονόμασαν Newt. Ένας γενετικός αλγόριθμος χρησιμοποιεί αλγοριθμικά ανάλογα των λειτουργιών γενετικής διασταύρωσης και μετάλλαξης για τη δημιουργία υποψηφίων προφίλ που κληρονομούν χρήσιμα χαρακτηριστικά από τους προγόνους τους, και χρησιμοποιούν τον ανταγωνισμό για την αναγνώριση και τη διατήρηση των καλύτερων από αυτά. Ο τελεστής διασταύρωσης εφαρμόζεται περιοδικά για να συνδυάσει τμήματα των δύο υποψηφίων προφίλ, τα οποία είναι μεταξύ εκείνων που παρήγαγαν τις υψηλότερες βαθολογίες (χρησιμοποιώντας ένα μέτρο ομοιότητας συνημιτόνου) για είδη που ο χρήστης αργότερα προσδιόριζε ως επιθυμητά. Ένας τελεστής μετάλλαξης εφαρμόζονταν μερικές φορές στο ισχύον όνομα της ομάδας συζήτησης, για να διερευνήσει αν τα υπάρχοντα υποψήφια προφίλ θα απέδιδαν καλά και σε ομάδες συζήτησης με παρόμοια ονόματα. Όλα τα υποψήφια προφίλ συνέβαλαν στην κατάταξη των εγγράφων που εμφανίζονταν στο χρήστη, αν και αυτά που απέδιδαν καλά με συνέπεια, συνέβαλαν πιο έντονα στην κατάταξη. Ως εκ τούτου, το ίδιο το προφίλ καθορίζονταν από τον πληθυσμό των υποψηφίων προφίλ, αντί για κάθε ένα μεμονωμένο υποψήφιο.

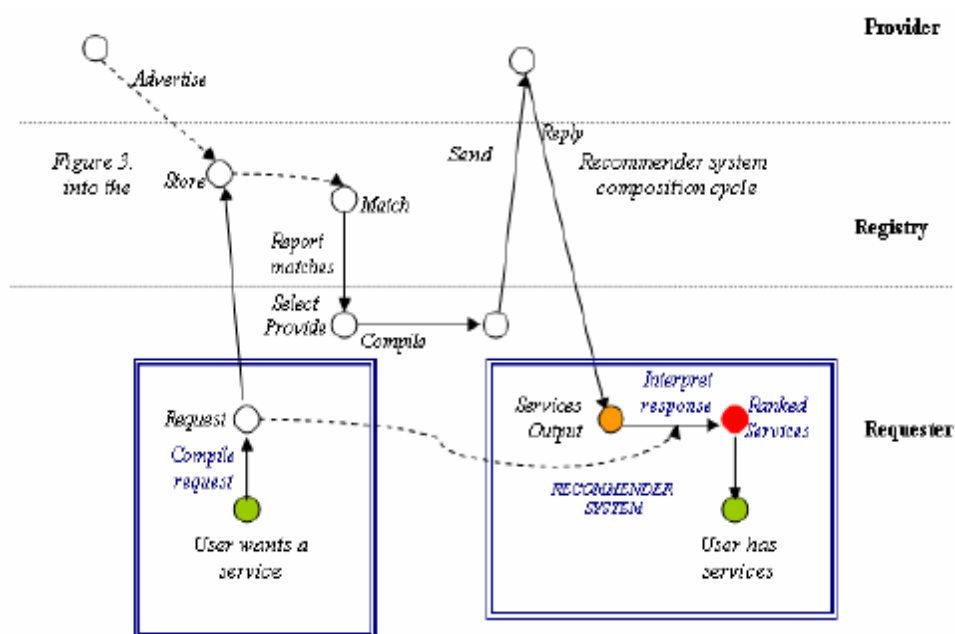
Μια παρόμοια προσέγγιση εφαρμόστηκε στην Amalthea (Moukas, 1997) δημιουργώντας ένα τεχνητό οικοσύστημα εξελισσόμενων πρακτόρων, οι οποίοι συνεργάζονται και ανταγωνίζονται σε ένα περιβάλλον περιορισμένων πόρων. Νέοι πράκτορες δημιουργούνται από διασταύρωση ή μετάλλαξη (ή και τα δύο). Και οι δύο τελεστές εφαρμόζονται στο εξελιζόμενο μέρος των πρακτόρων, τον γονότυπο. Το άλλο μέρος των πρακτόρων, ο φαινότυπος, περιέχει πληροφορία που δεν θα πρέπει να εξελιχθεί, Συστήματα Σύστασης βασισμένα σε Τεχνικές Νοημοσύνης Σμήνους

συνήθως οδηγίες σχετικά με τον χειρισμό του εξελίξιμου τμήματος. Ο τελεστής διασταύρωσης δύο σημείων λειτουργεί ως εξής: με δεδομένους δύο πράκτορες, επιστρέφει δύο νέους πράκτορες που κληρονομούν μέρος των διανυσμάτων λέξης-κλειδί των γονέων. Ο τελεστής επιλέγει τυχαία δύο σημεία στο διάνυσμα της λέξης-κλειδί και ανταλλάσσει όλα τα πεδία των δύο γονέων που βρίσκονται μεταξύ αυτών των σημείων, δημιουργώντας δύο νέους πράκτορες. Η μετάλλαξη είναι μια άλλη μέθοδος για τη δημιουργία απόγονων πρακτόρων. Ο τελεστής μετάλλαξης λαμβάνει τον γενότυπο ενός πράκτορα ως όρισμα και δημιουργεί έναν νέο πράκτορα που είναι μια τυχαία τροποποιημένη έκδοση των γονέων του. Οι συντελεστές στάθμισης από τις μεταλλαγμένες λέξεις-κλειδιά τροποποιούνται τυχαία, ενώ η νέα μεταλλαγμένη λέξη-κλειδί είναι μια τυχαία επιλεγμένη λέξη-κλειδί από έναν πράκτορα που ανήκει σε άλλη συστάδα.

Ωστόσο, το μέγεθος και η πολυπλοκότητα των δεδομένων αυξάνει με το συντακτικό επίπεδο στις πρόσφατες εξελίξεις στο web. Ανάμεσα στους σημαντικότερους πόρους του Web είναι εκείνοι που παρέχουν υπηρεσίες. Με τον όρο «υπηρεσία» εννοούμε Web τόπους που δεν παρέχουν απλώς στατικές πληροφορίες, αλλά επιτρέπουν σε κάποιον να εκτελέσει κάποια ενέργεια, όπως η πώληση ενός προϊόντος ή ο έλεγχος μιας φυσικής συσκευής. Ο σημασιολογικός ιστός (Semantic Web) θα πρέπει να επιτρέπει τους χρήστες να εντοπίζουν, να επιλέγουν, να απασχολούνται, να συνθέτουν και να παρακολουθούν τις υπηρεσίες που βασίζονται στο Web αυτόματα.

Ο γενικός όρος υπηρεσία Web (Web Service) στην πραγματικότητα δεν περιγράφει μια συνεκτική ή αναγκαστικά συνεπή έννοια, μάλλον, εμφανίζεται σαν ένα νέο πρότυπο για το Web καλύπτοντας το σύνολο των τεχνολογιών, των αρχιτεκτονικών, των πτυχών σε διαφορετικά επίπεδα της ηλεκτρονικής αγοράς (the e-Market). Χρησιμοποιείται πολύ συχνά χαλαρά για να δηλώσει ένα σύνολο σχετικών τεχνολογιών, οι οποίες περιλαμβάνουν: SOAP (Simple Object Access Protocol), Web Services Description Language (WSDL), OWL-S (Ontology Web Language Service), Universal Description, Discovery and Integration (UDDI), κλπ. Ο όρος υπηρεσία Web καθορίζει οποιαδήποτε περίπτωση όπου οποιοδήποτε λογισμικό προσφέρει πληροφορίες σε άλλο λογισμικό (Gil, 2004).

Αυτό σημαίνει μοντελοποίηση και αποθήκευση του μοντέλου του χρήστη και της πληροφορίας των μεταδεδομένων των χαρακτηριστικών του περιεχομένου χρησιμοποιώντας τυποποιημένες προδιαγραφές με δομές μεταδεδομένων βασισμένες σε XML. Νέα εργαλεία ανάλυσης μπορεί να αποδειχθούν ανεπαρκή και απαιτούνται πιο έξυπνες τεχνικές σε συνδυασμό με την επεξεργασία των μεταδεδομένων κατά τη διαδικασία σύστασης που εφαρμόζεται στη σύσταση υπηρεσιών σημασιολογικού ιστού (Web Semantic Services recommendation), όπως συνοψίζεται στο σχήμα που ακολουθεί.



Σχήμα 3: Σύστημα σύστασης βασισμένο στις υπηρεσίες σημασιολογικού ιστού

3.3 Recommenders στο e-learning

Η αύξηση του αριθμού των πηγών e-learning σημαίνει ότι χρειάζονται δυναμικές online εκπαιδευτικές υποδομές για την αποτελεσματική διαχείριση όλων των εκπαιδευτικών υπηρεσιών και στοιχείων. Τα e-learning συστήματα σύστασης θα συνιστούσαν ένα στοιχείο μάθησης σε έναν εκπαιδευόμενο με βάση τα καθήκοντα που έχει ήδη κάνει ο μαθητής και τις επιτυχίες του, και με βάση τις εργασίες που πραγματοποιήσαν άλλοι «παρόμοιοι» εκπαιδευόμενοι.

Η ομοιότητα των σπουδαστών θα μπορούσε να υπάρξει υιοθετώντας προφίλ χρηστών, ή θα μπορούσε να βασίζεται σε κοινά προηγούμενα μοτίβα πρόσβασης όπως ισχύει για το ηλεκτρονικό εμπόριο, αλλά τώρα το πλαίσιο είναι αρκετά διαφορετικό, λόγω της εμφάνισης ενός νέου ρόλου, του εκπαιδευτή. Ο στόχος κάθε διαδικασίας μάθησης είναι η απόκτηση ορισμένου περιεχομένου και η διαδικασία είναι συνθετότερη από την απλή αγορά ενός βιβλίου.

Σαν συνέπεια του σημασιολογικού ιστού, μια σημαντική συνεισφορά από τον επιστήμη υπολογιστών στη διαχείριση της γνώσης και στα e-learning συστήματα είναι η έννοια του αντικείμενου μάθησης (learning object - LO). Αυτό το στοιχείο έχει χαρακτηριστικά των ανεξάρτητων μονάδων, που είναι σε θέση να ξαναχρησιμοποιηθούν για άλλες εκπαιδευτικές καταστάσεις και πλατφόρμες. Κάθε ένα από τα LOs έχει μεταδεδομένα (δεδομένα σχετικά με τα δεδομένα) για την περιγραφή και τη διαχείρισή τους. Με αυτόν τον τρόπο είναι δυνατόν να γνωρίζουμε με τι είδους LO είμαστε αντιμέτωποι. Τα LOs χαρακτηρίζονται από το διαχωρισμό του περιεχομένου τους και της παρουσιάσής τους, και γι' αυτό το λόγο ένα σημαντικό θέμα που πρέπει να ληφθεί υπόψη είναι η αξιολόγηση και το φιλτράρισμά τους, προκειμένου να εξαχθεί μια σύσταση με βάση την πληροφορία των μεταδεδομένων τους. Μεταδεδομένα με βάση προδιαγραφές IMS, IMS LOM (IMS LOM, 2006), παρέχουν LOs με πληροφορίες σχετικά με την περιγραφή και τη διαχείρισή τους, ως εκ τούτου, είναι δυνατό να γνωρίζει κανείς αν τα χαρακτηριστικά τους είναι κατάλληλα για άλλες εκπαιδευτικές καταστάσεις. Η προδιαγραφή περιλαμβάνει δηλώσεις προσαρμογής για τον τρόπο που τα έγγραφα μεταδεδομένων πρέπει να είναι

οργανωμένα και για το πώς οι αιτήσεις πρέπει να συμπεριφέρονται προκειμένου να θεωρείται ότι ανταποκρίνονται στις απαιτήσεις LOM.

Σύμφωνα με αυτά, η διαχείριση της γνώσης για το e-learning βάσει των επαναχρησιμοποιούμενων μονάδων μάθησης σημαίνει τη δυνατότητα πρόσβασης σε συγκεκριμένο περιεχόμενο σύμφωνα με τις ανάγκες των διδασκομένων (Morales et al, 2006). Αυτό το στάδιο είναι δυνατό λόγω των προτύπων τα οποία καθιερώθηκαν σαν μια προσπάθεια να αποφευχθούν τα προβλήματα διαλειτουργικότητας στις διάφορες πλατφόρμες. Προκειμένου να ανταποκριθούν σ' αυτές τις απαιτήσεις αυτά τα στοιχεία θα πρέπει να πληρούν ένα ευρύ φάσμα απαιτήσεων συμπεριλαμβανομένων της εξατομίκευσης και της προσαρμογής. Όλα τα χαρακτηριστικά αυτά είναι γνωστά στους Recommenders EC.

Προκειμένου να υποστηριχθεί η διαδικασία διδασκαλίας και μάθησης μέσω συστημάτων e-learning υπάρχουν πολλές δυνατότητες συστημάτων διαχείρισης γνώσης, όπως η παράδοση και η αξιολόγηση των μαθημάτων, κλπ. ((Rosenberg, 2001), (Avgeriou, 2003)). Ωστόσο, σύμφωνα με τα LOs και τις δυνατότητες των προτύπων, είναι αναγκαίο να εξεταστεί ο τρόπος διαχείρισης ποιοτικών LOs, λαμβάνοντας υπόψη τα χαρακτηριστικά τους (βλέπε πίνακα 1).

Είναι γεγονός η μεγάλη σημασία των recommenders συστημάτων στο E-Learning, ενώ πρόκειται να υπάρξει μεγάλη ανάπτυξη σ' αυτά τα εργαλεία μέσω της επέκτασης του σημασιολογικού Ιστού.

Metadata Categories	Metadata Elements
1.General	1.2 Title 1.4 Description 1.5 Keywords 1.6 Coverage
5.Educational	5.1 Interactivity Type 5.2 Learning Resource Type 5.3 Interactivity Level 5.4 Semantic Density 5.6 Context 5.7 Typical Age Range 5.8 Difficulty 5.9 Typical Learning Time, 5.10 Description 5.11 Language
7.Relation	7.1 Type 7.2 Resource
8.Annotation	8.3 Description
9.Classification	9.1 Purpose 9.2 Taxon Path 9.2.1 Source 9.2.2 Taxon 9.2.2.1 Id 9.2.2.2 Entry 9.3 Description 9.4 Keyword

Πίνακας 1: Προτεινόμενες κατηγορίες και στοιχεία μεταδεδομένων για τη διαχείριση των αντικειμένων μάθησης LO

3.3.1 Σύστημα πολλών πρακτόρων στην εφαρμογή σύνθετων συστημάτων

Ένα σύνθετο σύστημα αποτελείται από έναν μεγάλο αριθμό αλληλεπιδρώντων μονάδων, οι οποίες όταν μελετώνται με μια σφαιρική προοπτική θα μπορούσε να θεωρηθεί ότι διαθέτουν σημαντικά χαρακτηριστικά γνωρίσματα και αναλογίες. Επιπλέον, ένα σύνθετο σύστημα είναι εγγενώς στοχαστικό στο εκτεταμένο χωροχρονικό σύμπαν του και ως εκ τούτου, κάποια από τα χαρακτηριστικά γνωρίσματά του θα μπορούσαν να φανερώνουν πρότυπα που εμφανίζονται συχνότερα από άλλα. Αυτές οι εμφανίσεις μετατρέπονται σε πρότυπα καθορίζοντας τις επαναλαμβανόμενες πτυχές του σύνθετου συστήματος και την εμφάνιση του φαινομένου.

Προκειμένου να περιγραφεί η σχέση ανάμεσα στο επί μέρους δυναμικό της κάθε μονάδας που γίνεται συλλογικό δυναμικό όταν οι μονάδες αλληλεπιδρούν, χρησιμοποιείται ο όρος αυτοοργάνωση (self-organization) ((Biebricher et al., 1995), (Bonabeau κ.ά., 1999)). Το φαινόμενο της αυτοοργάνωσης προκύπτει σαν ένας αυθόρμητος σχηματισμός μέσω της εξέλιξης και της διαφοροποίησης των δομών σύνθετης τάξης που σχηματίζονται σε μη γραμμικά δυναμικά συστήματα μέσω μηχανισμών ανατροφοδότησης που περιλαμβάνουν τα στοιχεία των συστημάτων. Αλλά είναι μακράν πιο σημαντικό ότι τα αλληλεπιδρώντα στοιχεία στα πολύπλοκα συστήματα επιφέρουν την συνολική δομή χρησιμοποιώντας μόνο τοπικές πληροφορίες, χωρίς αναφορά στο συνολικό πρότυπο που διαμορφώνεται. Για τη δημιουργία και τη μελέτη τέτοιου είδους πολύπλοκων συστημάτων, χρησιμοποιούνται συστήματα πολλών πρακτόρων (multi-agent systems - MAS), τα οποία αντιμετωπίζουν τις πτυχές της συνεργασίας, του σχηματισμού συνασπισμού και ορισμένα άλλα χαρακτηριστικά που ταιριάζουν με την περιγραφή των πολύπλοκων συστημάτων. Κάθε πράκτορας έχει ελλιπείς πληροφορίες ή δυνατότητες, κανένα συνολικό έλεγχο του συστήματος, αποκεντρωμένα δεδομένα, ικανότητα ασύγχρονων υπολογισμών και κοινωνική δυνατότητα.

Πολύ γενικά, τα στοιχεία του συστήματος αντιμετωπίζονται σαν multi-agents, σχετικά αυτόνομες οντότητες που έχουν ένα σύνολο από διαφορετικούς κανόνες για την αλληλεπίδραση μεταξύ τους. Οι κανόνες επικοινωνίας μπορούν επίσης να συσχετιστούν με τοπικές μεταβλητές, μειώνοντας την άμεση επικοινωνία μεταξύ των πρακτόρων, η οποία με τη σειρά της πρέπει να επηρεάζεται έντονα από τις περιβαλλοντικές αλλαγές με την απαραίτητη ευελιξία και διαπερατότητα. Αλλάζοντας τους κανόνες της αλληλεπίδρασης ή την επιρροή του περιβάλλοντος κατά τη διάρκεια της προσομοίωσης, κανείς μπορεί να είναι σε θέση να παρατηρήσει διαφορετικά είδη συλλογικής δυναμικής και της εμφάνισης των νέων ιδιοτήτων του συστήματος που δεν προβλέπονται εύκολα από τις βασικές εξισώσεις.

Διαφορετικές παραλλαγές των multi-agent μοντέλων εφαρμόζονται για την προσομοίωση κοινωνικο-οικονομικών διαδικασιών, τον σχηματισμό ανθρώπινων δομών, τις μεταφορές και τη βιομηχανία, κ.λπ., μοντέλα που κυμαίνονται από την οικολογία έως την τεχνητή ζωή. Προκειμένου να παράσχουμε στους πράκτορες των σύνθετων συστημάτων έναν μηχανισμό επικοινωνίας, δίνονται στο σύστημα ορισμένα χαρακτηριστικά βασισμένα στις φυσικές ιδιότητες ((Pagunak κ.ά., 2001), (Shehory κ.ά., 1999)) είτε προσαρμόζοντας τις φυσικές αρχές στη Distributed Artificial Intelligence, είτε με την εφαρμογή μοντέλων οργάνωσης που έχουν εξαχθεί από τη βιολογία (Bonabeau κ.ά., 1999).

3.3.2 Biologic Oriented Agentification

Τα πολυάριθμα είδη κοινωνικών εντόμων και το πιο γνωστό παράδειγμα, οι αποικίες μυρμηγκιών είναι η έμπνευση για τα οργανωτικά μοντέλα των σύνθετων συστημάτων (Dorigo και Stützle, 2004). Γενικά, το σμήνος πρέπει να πραγματοποιεί μια συλλογική εργασία: κάθε έντομο αποθέτει μια μικρή ποσότητα χημικών ουσιών στο π'ερασμά του, που ονομάζονται φερομόνες, επιτρέποντάς του να σημάνει έτσι τη διεύευσή του (προσθέτοντας μνήμη στη διαδρομή που διανύει), ενώ δίνει στα ομοειδή έντομα ορισμένες πληροφορίες σχετικά με το περιβάλλον και πληροφορίες σχετικά με τη δική του κατάσταση (επίτευξη επικοινωνίας). Δύο άτομα αλληλεπιδρούν έμμεσα μέσω του περιβάλλοντός τους. Στην πραγματικότητα, οι φερομόνες οδηγούν άμεσα σε μια συγκεκριμένη συμπεριφορά το άτομο που τις αντιλαμβάνεται, αυτό

το φαινόμενο ονομάζεται stigmergy. Σύμφωνα με τον Ramos (Ramos και Ajith, 2004), η stigmergy θα μπορούσε να οριστεί ως μια τυπική περίπτωση περιβαλλοντικής συνέργειας του να μαθαίνεις μέσω του περιβάλλοντος.

Οι φερομόνες λειτουργούν σαν χημικοί διαβιβαστές, επιτρέποντας στα μυρμήγκια να επικοινωνήσουν μεταξύ τους σε μια σύντομη απόσταση. Τα μυρμήγκια είναι ικανά για εξωτερική αποθήκευση των πληροφοριών στο περιβάλλον, επιτυγχάνοντας ένα είδος μνήμης.

Προκειμένου να γίνει κατανοητή η συλλογική συμπεριφορά, χρησιμοποιούνται προσομοιώσεις στον υπολογιστή για την εξέταση των παραμέτρων και των αλληλεπιδράσεών τους. Σε γενικές γραμμές, αυτές οι αποικίες εντόμων στον πραγματικό κόσμο παρέχουν τρεις λειτουργίες στις χημικές φερομόνες που υποστηρίζουν σκόπιμες ενέργειες στα μοντέλα πρακτόρων. Αυτές συγκεντρώνουν αποθέσεις φερομονών από επιμέρους πράκτορες, τις εξατμίζουν με την πάροδο του χρόνου (με αποτέλεσμα την αποφυγή υπερφόρτωσης και ξεχνώντας τις ξεπερασμένες πληροφορίες), και τις διαχέουν σε γειτονικές θέσεις (με αποτέλεσμα την παροχή μιας διαβάθμισης που να μπορούν να ακολουθήσουν οι πράκτορες). Σε ένα τέτοιο οικοσύστημα μπορούν να προσδιοριστούν οι παρακάτω ιδιότητες πρακτόρων:

- Αυτόνομη οντότητα. Κάθε πράκτορας λειτουργεί ανεξάρτητα και ασύγχρονα για να ικανοποιήσει τον στόχο του. Αυτό συνεπάγεται κατανομή σε επιμέρους μικρότερες λειτουργίες.
- Ικανός να ενεργήσει στο περιβάλλον του. Η βασική αλληλεπίδραση μεταξύ του πράκτορα και του περιβάλλοντος μπορεί να θεωρηθεί ως έμμεση χημική επικοινωνία που μεσολαβεί από μια εξωτερική αποθήκευση.
- Γνωρίζει το περιβάλλον του εν μέρει. Η γνώση αυτή βασίζεται στην αλληλεπίδραση του πράκτορα σε τοπικό ή μικροσκοπικό επίπεδο. Κάθε πράκτορας λειτουργεί μέσα σε οριοθετημένο ορθολογιστικό πλαίσιο και προσαρμόζεται συνεχώς.
- Εργάζεται για μεμονωμένους στόχους.
- Είναι ικανός να αλληλεπιδρά με άλλους πράκτορες. Ο πράκτορας έχει κοινωνικές αλληλεπιδράσεις.
- Κατά την εξέλιξη του συστήματος, εξαιτίας της αλληλεπίδρασης μεταξύ των πρακτόρων στο περιβάλλον, αναδύεται μια συλλογική συμπεριφορά που προκύπτει από προσαρμογή. Αυτή η συμπεριφορά παρατηρείται στο μακροσκοπικό επίπεδο. Έτσι εμφανίζονται ορισμένα multi-agent χαρακτηριστικά:
 - Η συνάθροιση πολυάριθμων πρακτόρων οφείλεται σε κοινά χαρακτηριστικά που ανακαλύπτονται κατά τη διάρκεια της εξέλιξης των στόχων.
 - Το σύστημα multi-agent έχει έναν στόχο να πετύχει.
 - Κανένας πράκτορας δεν ελέγχει τη συνολική εργασία. Τα μυρμήγκια εκτελούν εντυπωσιακά κατορθώματα συντονισμού χωρίς άμεσο inter-agent έλεγχο.
 - Το περιβάλλον είναι δυναμικό ή/και μερικώς περιγγραμμένο.

Κατά μια γενικότερη άποψη, τα οικοσυστήματα είναι περίπλοκα βιολογικά συστήματα στα οποία ένα ουσιώδες χαρακτηριστικό είναι η προσαρμογή. Ορισμένα μαθηματικά μοντέλα των οικοσυστημάτων προσομοιώνουν μοντέλα ετερογενών πρακτόρων που εξελίσσονται σε ένα σύστημα, σύμφωνα με την προσαρμογή τους με κάποια πλευρά του οικοσυστήματος. Συνήθως αυτοί οι πράκτορες ανταγωνίζονται για πόρους ή συνεργάζονται για έναν κοινό στόχο.

3.4 Συστήματα Σύστασης για e-learning Εφαρμογές

Ο Dorigo ((Dorigo και Di Caro, 1999), (Dorigo και Stützle, 2004)) όρισε τον εμπνευσμένο από τη βιολογία υπολογιστικό αλγόριθμο ACO (Ant Colony Optimization). Αυτός ο αλγόριθμος αποτελείται από κοινωνικά μυρμήγκια που κατασκευάζουν δίκτυα μονοπατιών που συνδέουν τις φωλιές τους με τις διαθέσιμες πηγές τροφής. Μαθηματικά, αυτά τα δίκτυα αποτελούν ελάχιστα spanning trees. Έτσι,

ελαχιστοποιούν την ενέργεια που δαπανούν για να φέρουν τροφή στη φωλιά. Ωστόσο, η γενική έρευνα χρησιμοποιώντας τεχνητά μυρμήγκια τείνει να επιλύσει πιο δύσκολα ζητήματα, προσθέτοντας πολυπλοκότητα στη συμπεριφορά των μυρμηγκιών για την επίλυση συγκεκριμένων πεδίων προβλημάτων.

Η προσπάθεια που έχει γίνει για την εξαγωγή και τη μελέτη των προτύπων του χρήστη σχετικά με το Internet είναι σημαντική, και ο Ramos (Ramos και Ajith, 2004) πρότεινε μια ACLUSTER (Ant Colony Cluster) εφαρμογή για να συσταδοποιήσει τα πρότυπα χρήσης του Web προκειμένου να προβλέψει τον όγκο κυκλοφορίας στο Web. Οι Gil και García-Peñalvo (2008) πρότειναν την επίλυση του προβλήματος της σύστασης με τη σχεδίαση πρακτόρων που θα καλύπτουν πτυχές σχετικά με τους εκπαιδευόμενους και τα LOs. Αυτοί οι πράκτορες που αλληλεπιδρούν στο μικροσκοπικό επίπεδο θα προκύψουν σε μια τελική δομή μέσω αυτοοργάνωσης από μια bottom-top MAS αρχιτεκτονική, που περιγράφει τη λύση. Η διαχείριση των στοιχείων στο πρόβλημα της σύστασης έχει πτυχές αυτοοργάνωσης, οι οποίες θα μπορούσαν να θεωρηθούν σαν ένα μοντέλο οικοσυστήματος λόγω ορισμένων σχετικών πλευρών:

- Παρόμοια LO πληροφορία ή σχετικές τάσεις για τη δημιουργία ομάδων.
- Η πιο απαραίτητη πληροφορία σε σχέση με τις εκπαιδευτικές ανάγκες του χρήστη μεγαλώνει και εξελίσσεται ενώ η υπόλοιπη ξεχνιέται και εξαφανίζεται.
- Η διάδοση όλων των πληροφοριών στο περιβάλλον τους πραγματοποιείται μέσω διαδικασιών αλληλεπιδράσεων περιορισμένης εμβέλειας και ανταποκρίνεται στο φαινόμενο της διάχυσης με την πάροδο του χρόνου.

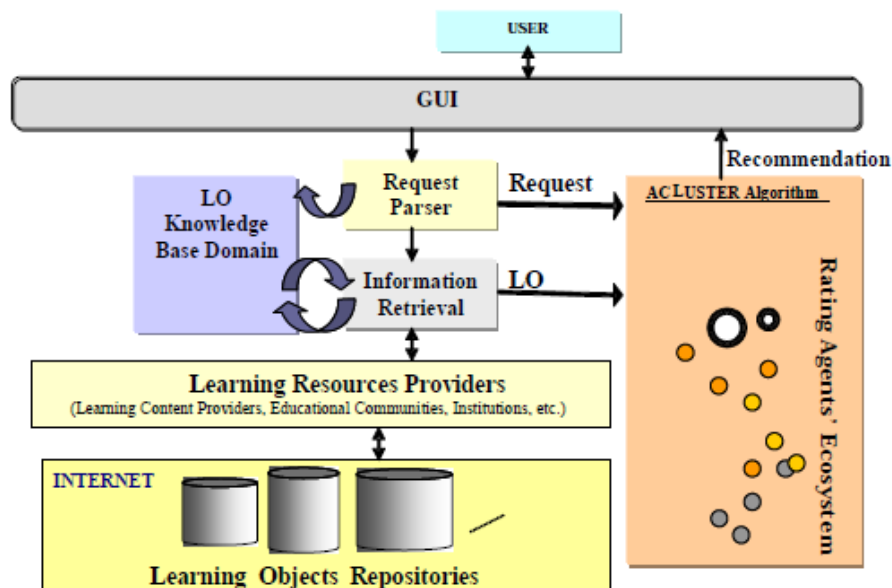
Με βάση τη φυσική επιλογή στο MAS μοντέλο, κάθε υπηρεσία αντιπροσωπεύεται από ένα μυρμήγκι, ενώ όλα τα μυρμήγκια συνεργάζονται και ανταγωνίζονται για να ικανοποιήσουν τις προσωπικές απαιτήσεις του χρήστη, ο οποίος τοποθετείται στο περιβάλλον επίσης σαν πηγή τροφής. Τόσο τα μυρμήγκια, όσο και οι πηγές τροφής είναι πράκτορες. Μόνο αυτά που περιγράφουν υπηρεσίες επαρκώς παρόμοιες ((Paolucci et al., 2002), (Paolucci et al., 2003)) με την αιτούμενη υπηρεσία, ταιριάζουν το αίτημα στο βαθμό που συμφωνούν με το προφίλ του χρήστη. Αυτά τα στοιχεία διαμορφώνουν τη βάση της αρχιτεκτονικής που προτείνεται στην επόμενη ενότητα.

3.4.1 Αρχιτεκτονική πρόταση

Η προτεινόμενη αρχιτεκτονική του συστήματος (βλέπε σχήμα 4) είναι καταμεμημένη σε τρία επίπεδα. Το πρώτο επίπεδο περιέχει το γραφικό περιβάλλον διεπαφής χρήστη (GUI) και διοχετεύει τα δεδομένα επικοινωνίας: ο χρήστης υποβάλει τα αιτήματά του, πλοηγείται, λαμβάνει και επιλέγει τις συστάσεις. Η αλληλεπίδραση με τον χρήστη έχει ισχυρές πλευρές που εξαρτώνται από τον τομέα (περιγράφεται σημασιολογικά).

Το δεύτερο επίπεδο περιλαμβάνει το κύριο επίπεδο της εφαρμογής όπου βρίσκονται οι στρατηγικές των συστάσεων. Αυτό είναι σχεδιασμένο σαν το οικοσύστημα ενός πράκτορα. Το οικοσύστημα είναι ο γενήτορας μιας δυναμικής παρουσίας του περιβάλλοντος στην χωροχρονική του εξέλιξη. Το οικοσύστημα που ορίζεται στο δεύτερο επίπεδο αποτελείται από ένα διακριτό περιβάλλον, όπου οι πράκτορες εγκυμονούν τα χαρακτηριστικά τομέα του χρήστη. Αυτοί οι πράκτορες λαμβάνουν και εκπέμπουν πληροφορίες μέσω του περιβάλλοντος.

Τέλος, το τρίτο επίπεδο, σχετίζεται με την πηγή πληροφοριών, περιέχει τις υπηρεσίες γνωσιακής βάσης τομέα (οντολογική γνώση) και πολλές επεκτάσεις στον παγκόσμιο ιστό που περιέχουν διαμεσολαβητές δεδομένων. Σε αυτό το επίπεδο είναι επίσης αποθηκευμένη και μια αποθήκη πληροφοριών για το ίδιο το σύστημα.



Διάγραμμα 4: Αρχιτεκτονική Συστήματος Recommender (Gil και García-Peñalvo 2008)

3.4.2 Ο αλγόριθμος συσταδοποίησης και ταξινόμησης μνημυγκιών

Ορισμένα είδη μυρμηγκιών τα *Pheidole pallidula*, τα *Laius niger* και τα *Messor sancta* σχηματίζουν σωρούς στοιχείων όπως είναι τα σώματα νεκρών μνημυγκιών, οι προνύμφες, ή οι κόκκοι άμμου. Αυτά τα μυρμηγκία αποθέτουν στοιχεία σε αρχικά τυχαίες θέσεις. Όταν τα άλλα μυρμηγκία αντιλαμβάνονται τα στοιχεία αυτά, υποκινούνται να αποθέσουν στοιχεία δίπλα σε αυτά, με το νεκροταφείο μνημυγκιών να αποτελεί μια οργάνωση συσταδοποίησης και η ταξινόμηση των προνυμφών, έναν τύπο αυτοοργάνωσης και προσαρμοσμένης συμπεριφοράς. Η συμπεριφορά συσταδοποίησης και ταξινόμησης των μνημυγκιών (Clustering and Sorting Ant Algorithm) έχει υποκινήσει ερευνητές να σχεδιάσουν νέους αλγόριθμους για την ανάλυση των δεδομένων. Αντικείμενα που τοποθετούνται δίπλα το ένα στο άλλο από τον αλγόριθμο ταξινόμησης έχουν παρόμοια χαρακτηριστικά.

Οι Deneubourg et al. (1991) πρότειναν ένα πρώτο μοντέλο (ονομάζεται BM, για το βασικό μοντέλο) όπου ένας πληθυσμός ant-like πρακτόρων μετακινούμενος τυχαία σε ένα διδιάστατο πλέγμα επιτρέπεται να μετακινήσει βασικά αντικείμενα, συσσωρεύοντας εκείνα του ίδιου τύπου και σχηματίζοντας συστάδες (Chrétien, 1996). Αυτός ο αλγόριθμος χρησιμοποιήθηκε στην εφαρμογή των ρομπότ.

Η πιθανότητα p_p ένας τυχαία κινούμενος, χωρίς φορτίο πράκτορας (που αντιπροσωπεύει ένα μυρμηγκί στο μοντέλο) να πάρει ένα στοιχείο, δίνεται από την εξίσωση 1, όπου f είναι το αναμενόμενο κλάσμα των στοιχείων στην γειτονιά του πράκτορα, και το k_1 είναι σταθερό. Με τον ίδιο τρόπο η πιθανότητα p_d ένας τυχαία κινούμενος, με φορτίο πράκτορας θα αποθέσει ένα στοιχείο είναι η εξίσωση 2, όπου το k_2 είναι σταθερό.

$$p_p = \left(\frac{k_1}{k_1 + f} \right)^2 \quad \text{Where} \quad \begin{array}{l} \text{if } f \ll k_1 \quad p_p \rightarrow 1 \\ \text{if } f \gg k_1 \quad p_p \rightarrow 0 \end{array}$$

Εξίσωση 1: Πιθανότητα BM για τη λήψη ενός στοιχείου

$$p_d = \left(\frac{f}{k_2 + f} \right)^2$$

Εξίσωση 2: Πιθανότητα BM για την απόθεση ενός στοιχείου

Αυτή η μέθοδος στη συνέχεια γενικεύθηκε από τους Lumer και Faieta (στο εξής ο αλγόριθμος LF) (Lumer και Faieta, 1995), ο οποίος εφαρμόστηκε στη διερευνητική ανάλυση δεδομένων. Έδειξαν ότι το μοντέλο τους παρέχει έναν τρόπο διερεύνησης διαστημάτων σύνθετων πληροφοριών, όπως τα έγγραφα ή οι σχεσιακές βάσεις δεδομένων. Ο LF ορίζει μια απόσταση ή ανομοιότητα μεταξύ αντικειμένων στο χώρο των χαρακτηριστικών του αντικειμένου. Τα αντικείμενα μπορούν να περιγραφούν από ένα πεπερασμένο αριθμό χαρακτηριστικών που σχετίζονται με την τιμή, επιτρέποντας την πρόσβαση στις πληροφορίες και συγκρίνοντάς τις σε ένα χώρο n-διαστάσεων, εξ' ου και η δυνατότητα απόκτησης διαφορετικών συστάδων. Ο LF αλγόριθμος λειτουργεί ως εξής:

Εστω ότι $d(o_i, o_j)$ είναι η απόσταση μεταξύ δύο αντικειμένων στο χώρο των χαρακτηριστικών. Ας υποθέσουμε ότι είναι ένας πράκτορας βρίσκεται στην τοποθεσία r τη χρονική στιγμή t και βρίσκει ένα αντικείμενο o_i σ' αυτή την τοποθεσία. Η τοπική πυκνότητα $f(o_i)$ θα σε σχέση με το αντικείμενο o_i στην τοποθεσία r δίνεται για $f > 0$ από την εξίσωση:

$$f(o_i) = \begin{cases} \frac{1}{s^2} \sum_{o_j \in \text{Neigh}_{(sxs)}(r)} \left[1 - \frac{d(o_i, o_j)}{\alpha} \right] & \text{if } f > 0 \\ 0 & \text{otherwise} \end{cases}$$

Εξίσωση 3: LF τοπική πυκνότητα σε σχέση με το αντικείμενο o_i στην τοποθεσία r

Το $f(o_i)$ είναι ένα μέτρο της μέσης ομοιότητας του αντικειμένου o_i με τα άλλα αντικείμενα o_j παρόντα στην γειτονιά του o_i . Το α είναι ένας παράγοντας που καθορίζει την κλίμακα για ανομοιότητα. Αυτός ο παράγοντας λειτουργεί ως ρυθμιστής της ομοιότητας που επιτρέπεται μεταξύ των διαφόρων στοιχείων για να σχηματίσουν την ίδια συστάδα ή όχι. Οι Lumer και Faieta καθόρισαν πιθανότητες λήψης και ρίψης (Εξίσωση 4, Εξίσωση 5) όπου k_1 και k_2 είναι δύο σταθερές παρόμοιες με εκείνες του bm.

$$p_p(o_i) = \left(\frac{k_1}{k_1 + f(o_i)} \right)^2$$

Εξίσωση 4: LF πιθανότητα λήψης

$$p_d(o_i) = \begin{cases} 2f(o_i), & \text{if } f(o_i) < k_2 \\ 1, & \text{if } f(o_i) \geq k_2s \end{cases}$$

Εξίσωση 5: LF πιθανότητα ρίψης

Οι Lumer και Faieta, εφάρμοσαν τον αλγόριθμο σε μια βάση δεδομένων που περιείχε το προφίλ 1650 πελατών μιας Τράπεζας. Χαρακτηριστικά των προφίλ περιελάμβαναν την οικογενειακή κατάσταση, το φύλο, την κατοικία, την ηλικία, μια λίστα των τραπεζικών υπηρεσιών που χρησιμοποιούνται από τον πελάτη κ.λπ. Δεδομένης της ποικιλίας των χαρακτηριστικών, ορισμένα από τα οποία ποιοτικά, άλλα ποσοτικά, έπρεπε να ορίσουν πολλά μέτρα ανομοιότητας και να τα συνδυάσουν σε ένα συνολικό μέτρο ανομοιότητας.

Οι Lumer και Faieta εξέλιξαν το μοντέλο (Lumer και Faieta, 1994) με μερικά χαρακτηριστικά γνωρίσματα για να διορθώσουν την τάση δημιουργίας περισσότερων συστάδων από τις επιθυμητές. Αυτά τα τρία νέα χαρακτηριστικά τροφοδοτούν τους πράκτορες με διαφορετικές ταχύτητες κίνησης, μαζί με βραχυπρόθεσμη μνήμη (έτσι, οι πράκτορες μπορούν να θυμούνται τα τελευταία m στοιχεία που ρίφθηκαν) και το σύστημα είναι εξοπλισμένο με συμπεριφορικούς διακόπτες που ενεργοποιούν ορισμένες ενέργειες ή που έχουν την δυνατότητα να καταστρέψουν συστάδες. Παρά τα ενδιαφέροντα αποτελέσματα, δεν είναι προφανές ότι αυτός ο αλγόριθμος έχει μέλλον με όρους απόδοσης σε υπολογιστικό χρόνο.

Πρόσφατα, οι Ramos et al., σε διάφορες δημοσιεύσεις (Ramos κ.ά., 2002) (Ramos και Merelo, 2002) και Ramos και Ajith (Ramos και Ajith, 2004) έχουν επεκτείνει το μοντέλο των Deneubourg, Lumer και Faieta που προαναφέρθηκε. Αυτός ο αλγόριθμος, που ονομάζεται Ant Clustering Algorithm (ACLUSTR), αποφεύγει την πρόσθετη πολυπλοκότητα των προηγούμενων αλγόριθμων (βραχυπρόθεσμη μνήμη, πολλοί τύποι μυρμηγκιών που κινούνται με διαφορετικές ταχύτητες), εισάγοντας ίχνη φερομόνης για την επίτευξη μη επιτηρούμενης συσταδοποίησης.

Οι Ramos και Merelo χρησιμοποίησαν ένα επαναπροσδιορισμένο μοντέλο των Chialvo και Millonas (Chialvo και Millonas, 1995), όπου ένα μεμονωμένο μυρμήγκι μπορεί να περιγραφεί από τη θέση και την κατεύθυνσή του. Οι πιθανότητες ένα μυρμήγκι να μετακινηθεί από ένα ζεύγος θέσης και κατεύθυνσης σε οποιοδήποτε άλλο ζεύγος καθορίζονται από τη σταθμισμένη συνάρτηση φερομόνης.

$$W(\sigma) = \left(1 + \frac{\sigma}{1 + \delta\sigma}\right)^\beta$$

Εξίσωση 6: Πιθανότητα μετακίνησης σε μια θέση με πυκνότητα φερομόνης $\sigma(r)$

Όπου η τιμή β καθορίζει την τυχειότητα με την οποία ένα μυρμήγκι ακολουθεί το ίχνος της φερομόνης, $1/\delta$ είναι αισθητήρια ικανότητα του μυρμηγκιού, η οποία περιγράφει το γεγονός ότι η δυνατότητα του κάθε μυρμηγκιού να αντιλαμβάνεται την φερομόνη μειώνεται σε υψηλή περιεκτικότητα. Η κανονικοποιημένη πιθανότητα μετάβασης από τη θέση i ως τη θέση k δίνεται από:

$$P_{ik} = \frac{w(\sigma_i)w(\Delta_i)}{\sum_{j/k} w(\sigma_j)w(\Delta_j)}$$

Εξίσωση 7: Πιθανότητα μετάβασης από το κελί i στο κελί k

Όπου το Δ_j μετράει το μέγεθος της διαφοράς στον προσανατολισμό από την προηγούμενη κατεύθυνση τη χρονική στιγμή $t-1$. Κάθε άτομο αφήνει μια σταθερή ποσότητα φερομόνης η από το κελί

στο οποίο βρίσκεται σε κάθε βήμα t , και επίσης αυτή η φερομένη εξασθενεί σε κάθε χρονικό βήμα με ρυθμό k .

Οι δύο μεγάλοι παράγοντες που ασκούν επιρροή σε οποιαδήποτε τοπική ενέργεια των μυρμηγκιών είναι ο αριθμός των αντικειμένων στη γειτονιά τους και η ομοιότητά τους. Οι Ramos και Ajith (Ramos και Ajith, 2004) όρισαν τη συνάρτηση πιθανότητας για τη λήψη ή την ρίψη σαν μια συνάρτηση κινήτρων διαφορετικών εντάσεων (αριθμός στοιχείων και η ομοιότητά τους), στην τοποθεσία r :

$$P_p = (1 - \chi) \cdot \varepsilon \quad P_d = \chi \cdot \delta$$

Εξίσωση 8: Συναρτήσεις πιθανότητας για τη λήψη ή την ρίψη

Όπου το χ ορίζεται ως το κατώφλι απόκρισης που σχετίζεται με τον αριθμό των στοιχείων n , παρόν σε μια 3×3 περιοχή γύρω από το r , και d είναι η ομοιότητα μεταξύ των αντικειμένων σαν Ευκλείδεια κανονικοποιημένη απόσταση που υπολογίζεται σε όλα τα ζευγάρια των αντικειμένων που βρίσκονται στην 3×3 περιοχή γύρω από το r .

$$\chi = \frac{n^2}{n^2 + 5^2}$$

Εξίσωση 9: Κατώφλι απόκρισης που σχετίζεται με τον αριθμό των αντικειμένων στη γειτονιά

όπου τα δ και ε ορίζονται σαν συναρτήσεις του κατωφλίου απόκρισης που σχετίζονται με την ομοιότητα των αντικειμένων σε περιπτώσεις όπου κάποιο απορρίπτεται και λαμβάνεται αργότερα:

$$\delta = \left(\frac{k_1}{k_1 + d} \right)^2 \quad \varepsilon = \left(\frac{d}{d + k_2} \right)^2$$

Εξίσωση 10: Συναρτήσεις κατωφλίου απόκρισης

Ο αλγόριθμος ACLUSTER εφαρμόστηκε στη συσταδοποίηση εγγράφων κειμένου (Ramos και Merelo, 2002), (Ramos και Ajith, 2004) και εισήγαγε ένα νέο είδος εξόρυξης δεδομένων βασισμένο σε παραδείγματα stigmergy υβριδιοποιώντας την εμπνευσμένη από τη βιολογία Νοημοσύνη Σμήνους με Εξελκτικούς Αλγόριθμους. Αυτή η εργασία παρέχει ενθαρρυντικά αποτελέσματα.

3.4.3 Συμπεράσματα

Η μελέτη των Gil και García-Reñalvo (2008) έριξε μια σύντομη ματιά στην αυτοοργάνωση κατά τη διάρκεια εξέτασης ορισμένων πτυχών που συνδέονται με τα Συστήματα Σύστασης. Εξηγεί τη βάση στην οποία προτείνεται ένας μηχανισμός σύστασης στο e-learning βασισμένος σε σύνθετα συστήματα με έναν αλγόριθμο εμπνευσμένο από τη βιολογία. Το περιβάλλον περιέχει τη γνώση όλου του φαινομένου, των πραγματικών γεγονότων που αφορούν τον τομέα, τις παραστάσεις των προτιμήσεων του χρήστη και επίσης τα χαρακτηριστικά των LOs, μοντελοποιημένα και καθορισμένα χρησιμοποιώντας τις πληροφορίες μάθησης IMS ή την προδιαγραφή LOM. Από την αλληλεπίδραση, αυτά τα στοιχεία παράγουν την τελική αναπαράσταση στο macro επίπεδο, η οποία προσαρμόζει μια δυναμική και εξατομικευμένη παρουσίαση της σύστασης.

Ο στόχος είναι να έχουμε μια δομική και άμεση σύνδεση ανάμεσα στο σύνολο των LO στοιχείων που χρειάζεται ο χρήστης, στηριζόμενος στο σημασιολογικό περιεχόμενο και στην αξιολόγηση της ποιότητας, και στον χώρο των πρακτόρων που αντιπροσωπεύουν τη σύσταση. Εργαστήκαμε στον αλγόριθμο ACLUSTER για να την εισαγωγή νέων μετρήσεων που θα επιτρέπουν την αναγνώριση της ομοιότητας μεταξύ των LOs στο περιβάλλον. Αυτή η εργασία προσπάθησε να δημιουργήσει λύσεις για εφαρμογές πραγματικής ζωής, με βάση τη Νοημοσύνη Σμήνους, που δίνει μεγάλες υποσχέσεις για περαιτέρω πρόοδο στον τομέα αυτό.

ΚΕΦΑΛΑΙΟ 4

4. Particle Swarm Optimization Recommender System

4.1 ΕΙΣΑΓΩΓΗ

Η διαδικασία λήψης αποφάσεων αποτελεί αναπόσπαστο μέρος της καθημερινής ζωής. Όταν αντιμετωπίζουμε ένα δίλημμα, οι περισσότεροι από εμάς είναι πιθανό να συγκεντρώσουμε ορισμένες σχετικές πληροφορίες προτού πάρουμε μια τεκμηριωμένη απόφαση. Αυτές οι πληροφορίες μπορεί να προέρχονται από διάφορες πηγές – γνωστά γεγονότα, προκαθορισμένους κανόνες, ή γνώμες άλλων. Πολλές lifestyle δραστηριότητες, όπως οι αγορές ρούχων, το φαγητό έξω και ο κινηματογράφος εξαρτώνται κατά πολύ από τις προσωπικές προτιμήσεις. Είναι αυτή η "προσωπική" πτυχή που καθιστά πολύ πιο δύσκολο το να τα συστήσει κανείς, καθώς έχουν μια μεγαλύτερη εξάρτηση από την προσωπικότητα του κάθε ατόμου. Ως εκ τούτου, οι άνθρωποι συχνά επιδιώκουν συμβουλές από τους φίλους τους, που μπορούν να κάνουν προτάσεις με βάση τη σχέση τους και την οικειότητά τους. Ωστόσο, υπάρχουν στιγμές που οι φίλοι δεν μπορούν να βοηθήσουν εξαιτίας του μεγάλου αριθμού των διαθέσιμων επιλογών.

Τα Συστήματα σύστασης παρέχουν έναν τρόπο παράκαμψης αυτού του προβλήματος. Όπως υποδηλώνει το όνομά τους, η αποστολή τους είναι να προτείνουν προϊόντα στον πελάτη βάσει των προτιμήσεών του. Αυτά τα συστήματα χρησιμοποιούνται συχνά από E-commerce ιστοσελίδες ως εργαλεία marketing για την αύξηση των εσόδων παρουσιάζοντας τα προϊόντα που ο πελάτης είναι πιθανό να αγοράσει. Ένας δικτυακός τόπος που χρησιμοποιεί ένα σύστημα σύστασης μπορεί να εκμεταλλευτεί τη γνώση σχετικά με τις προτιμήσεις και τις αντιπάθειες των πελατών για την κατανόηση των επιμέρους αναγκών τους και ως εκ τούτου, να αυξήσει την αφοσίωση των πελατών του.

Παρακάτω θα περιγραφεί η χρήση του αλγόριθμου βελτιστοποίησης σμήνους σωματιδίων (PSO) για τη ρύθμιση ενός αλγόριθμου ταιριάσματος προφίλ (profile-matching) μέσα σε ένα recommender σύστημα, προσαρμόζοντάς τον στις προτιμήσεις των μεμονωμένων χρηστών Supriya Ujjin και Peter J. Bentley (2003). Αυτό επιτρέπει στο σύστημα σύστασης να κάνει πιο ακριβείς προβλέψεις των προτιμήσεων και των αντιπαθειών των χρηστών, και ως εκ τούτου καλύτερες συστάσεις για τους χρήστες. Η τεχνική PSO εφευρέθηκε από τους Eberhart και Kennedy το 1995 και εμπνεύστηκε από τις συμπεριφορές κοινωνικών ζώων, όπως τα πουλιά και τα ψάρια. Ο αλγόριθμος είναι απλός και περιλαμβάνει την προσαρμογή λίγων παραμέτρων. Με μικρή τροποποίηση, μπορεί να εφαρμοστεί σε ένα ευρύ φάσμα εφαρμογών. Για αυτόν το λόγο, ο PSO έχει αυξανόμενο ενδιαφέρον από τους ερευνητές σε διάφορους τομείς.

Από τη βιβλιογραφία, φαίνεται ότι ο ορισμός του όρου "σύστημα recommender" διαφέρει ανάλογα με τον συγγραφέα. Ορισμένοι ερευνητές χρησιμοποιούν τις έννοιες: "σύστημα recommender" ("recommender system"), "συνεργατική διοίκηση" ("collaborative filtering") και "κοινωνικό φιλτράρισμα" ("social filtering") ή συνεργατική διοίκηση εναλλάξ (Breese, J.S. et al 1998). Αντίθετα, άλλοι θεωρούν το "σύστημα recommender" ως μια γενική περιγραφή που αντιπροσωπεύει διάφορες τεχνικές σύστασης/πρόβλεψης, συμπεριλαμβανομένων της συνεργατικής διοίκησης, του κοινωνικού φιλτραρίσματος και του φιλτραρίσματος με βάση το περιεχόμενο, των Bayesian δικτύων και των κανόνων συσχέτισης (association rules) (Terveen, L. and Hill, 2001 και Delgado, J.A. February 2000).

Ο ιστότοπος MovieLens (<http://www.movielens.umn.edu>), είναι ένας γνωστός πειραματικός ιστότοπος σύστασης ταινιών, που κάνει χρήση της τεχνολογίας συνεργατικής διοίκησης για να κάνει τις προτάσεις του. Αυτή η τεχνολογία καταγράφει τις προτιμήσεις του χρήστη για να δημιουργήσει ένα προφίλ ζητώντας από το χρήστη να αξιολογήσει ταινίες. Αναζητά παρόμοια προφίλ (δηλ., χρήστες που

μοιράζονται τις ίδιες ή παρόμοιες προτιμήσεις) και τα χρησιμοποιεί για να δημιουργήσει νέες προτάσεις (το σύνολο των δεδομένων που συλλέγονται μέσω του MovieLens δικτυακού τόπου έχει καταστεί διαθέσιμο για τους σκοπούς της έρευνας και χρησιμοποιείται για τον έλεγχο του PSO recommender συστήματος.)

Αντίθετα, ο ιστότοπος LIBRA (<http://www.cs.utexas.edu/users/Libra>) συνδυάζει μια προσέγγιση με βάση το περιεχόμενο με μηχανική μάθηση για να προβεί σε συστάσεις βιβλίων. Η προσέγγιση με βάση το περιεχόμενο διαφέρει από τη συνεργατική διοίκηση, στο ότι αναλύει το περιεχόμενο των στοιχείων που συστήνονται. Επιπλέον, ο κάθε χρήστης αντιμετωπίζεται μεμονωμένα - δεν υπάρχει καμία έννοια "κοινότητας", που να αποτελεί τη βάση της συνεργατικής διοίκησης.

Ο ιστότοπος Dooyoo (<http://www.dooyoo.co.uk>) λειτουργεί με ελαφρώς διαφορετικό τρόπο. Κι αυτός είναι πολύ χρήσιμος ιστότοπος, ο οποίος παρέχει συστάσεις για αυτούς που αναζητούν συμβουλές, αλλά επικεντρώνεται κυρίως στο να συλλέγει ποιοτικές γνωμοδοτήσεις από τους χρήστες του, και στο να τις καθιστά στη συνέχεια διαθέσιμες σε άλλους. Οι επισκέπτες του ιστότοπου συχνά υποβάλλουν σχόλια σχετικά με προϊόντα ή υπηρεσίες που κυμαίνονται από ιαματικά spa μέχρι κινητά τηλέφωνα. Αυτά τα στοιχεία είναι κατηγοριοποιημένα και παρουσιασμένα με παρόμοιο τρόπο όπως σε μια δομημένη μηχανή αναζήτησης, σαν το Yahoo!

Ερευνητές του University of the West of England επίσης εργάζονται για ένα Σύστημα σύστασης ταινιών (Cayzer, S. and Aickelin U. 2001). Η ιδέα τους είναι να χρησιμοποιήσουν το ανοσοποιητικό σύστημα για την αντιμετώπιση του προβλήματος του ταιριάσματος προτιμήσεων και συστάσεων. Οι προτιμήσεις των χρηστών αντιμετωπίζονται σαν ομάδα αντισωμάτων και ο ενεργός χρήστης σαν το αντιγόνο. Η διαφορά στην προσέγγισή τους από τις άλλες υπάρχουσες μεθόδους είναι ότι δεν ενδιαφέρονται να βρουν το καλύτερο ταίριασμα, αλλά ένα σύνολο ποικίλων αντισωμάτων που να αποτελούν ένα κοντινό ταίριασμα.

Το Σύστημα σύστασης γενετικού αλγόριθμου (GA) που χρησιμοποιείται στα πειράματα περιγράφεται στο άρθρο του Ujjiin, S. (2002). Το σύστημα ακολουθεί την ίδια δομή με το σύστημα recommender PSO, που συζητείται αργότερα στο τμήμα III. Η μόνη διαφορά είναι ότι το σύστημα GA χρησιμοποιεί έναν γενετικό αλγόριθμο για να διαμορφώσει τα βάρη των χαρακτηριστικών για τον ενεργό χρήστη (στο παρόν έγγραφο ένας αλγόριθμος PSO χρησιμοποιείται για τον ίδιο σκοπό). Ένα άτομο στον πληθυσμό αντιπροσωπεύει μια πιθανή λύση, η οποία στην προκειμένη περίπτωση ήταν ένα σύνολο των βαρών των χαρακτηριστικών που ορίζει την προτίμηση του ενεργού χρήστη. Εφαρμόστηκε μια συνάρτηση καταλληλότητας (fitness function) χρησιμοποιώντας μια τροποποιημένη συνάρτηση Ευκλείδειας απόστασης. Τα πειραματικά αποτελέσματα έδειξαν ότι το σύστημα GA απέδωσε πολύ καλά σε σύγκριση με μια μη-προσαρμοσμένη προσέγγιση βασισμένη στον αλγόριθμο Pearson.

4.2 Βελτιστοποίηση σμήνους σωματιδίων

Αν και η βελτιστοποίηση σμήνους σωματιδίων είναι μια εξελικτική τεχνική βασισμένη στον πληθυσμό όπως οι γενετικοί αλγόριθμοι, διαφέρει στο ότι κάθε σωματίδιο ή λύση περιέχει θέση, ταχύτητα και επιτάχυνση. Η ταχύτητα και η επιτάχυνση είναι υπεύθυνες για την αλλαγή της θέσης του σωματιδίου προκειμένου να εξερευνηθεί τον χώρο για όλες τις πιθανές λύσεις, αντί να χρησιμοποιήσει υπάρχουσες λύσεις για την αναπαραγωγή (Blackwell, T.M. and Bentley, P.J. 2002.). Καθώς τα σωματίδια μετακινούνται στο χώρο, δειγματοληπτούν διαφορετικές θέσεις. Κάθε θέση έχει μια τιμή καταλληλότητας σύμφωνα με το πόσο καλή είναι στο να ικανοποιήσει το στόχο, σε αυτήν την περίπτωση, να ορίσει τις προτιμήσεις του χρήστη. Λόγω των κανόνων που διέπουν τη διαδικασία σχηματισμού σμήνους, τα σωματίδια τελικά θα δημιουργήσουν σμήνος γύρω από την περιοχή στο χώρο όπου υπάρχουν οι καταλληλότερες λύσεις. Ο PSO αλγόριθμος είναι δημοφιλής για πολλές διαφορετικές εφαρμογές, για παράδειγμα:

Οι Van den Bergh et. Al εφάρμοσαν την αρχική τεχνική Βελτιστοποίησης σμήνους σωματιδίων, καθώς και μια παραλλαγή της τον Cooperative Particle Swarm Optimiser για να εκπαιδεύσουν Νευρωνικά δίκτυα (van den Bergh, F. και Engelbrecht, A.P. 2001). Σμήνη σωματιδίων χρησιμοποιήθηκαν για να βρεθούν τα βέλτιστα βάρη ενός Product Unit Neural Network. Ο Cooperative Optimiser χρησιμοποιεί πολλαπλά σμήνη, όπου κάθε σμήνος χειρίζεται ένα τμήμα του διανύσματος που βελτιστοποιείται. Ο συντελεστής διαχωρισμού (split factor), μια παράμετρος που προστίθεται στον αρχικό αλγόριθμο PSO, καθορίζει τον αριθμό των συστατικών μερών για κάθε σωματίδιο, που είναι επίσης ο αριθμός των σμηνών που χρησιμοποιείται.

Το Swarm Music είναι ένας αλληλεπιδραστικός αυτοσχεδιαστής μουσικής από τους Blackwell και Bentley που χρησιμοποιεί ένα σμήνος σωματιδίων τα οποία ερμηνεύονται σαν μουσικά γεγονότα. Αυτά τα σωματίδια αλληλεπιδρούν μεταξύ τους σύμφωνα με τους κανόνες που βασίζονται σε μοντέλα σμήνους. Ο χώρος της μουσικής έχει διαστάσεις που αντιπροσωπεύουν μουσικές παραμέτρους όπως ο παλμός, ο τόνος και η ένταση. Το σμήνος έλκεται προς τους στόχους που είναι εξωτερικά μουσικά γεγονότα που καταγράφονται και τοποθετούνται στο χώρο. Καθώς τα σωματίδια μετακινούνται γύρω από το χώρο της μουσικής των μεταβαλλόμενων παραμέτρων, παράγονται αυτοσχεδιασμοί αλληλεπιδραστικά με εξωτερικούς μουσικούς (Blackwell, T.M. and Bentley, P.J. 2002).

4.3 Επισκόπηση του Συστήματος

Το σύστημα που περιγράφεται εδώ, βασίζεται σε μια προσέγγιση συνεργατικής διοίκησης, δημιουργώντας προφίλ των χρηστών και στη συνέχεια, χρησιμοποιώντας έναν αλγόριθμο για να βρεί προφίλ που να μοιάζουν μ' αυτό του τρέχοντα χρήστη. (αναφερόμαστε στον τρέχοντα χρήστη ως ενεργός χρήστης, A). Επιλεγμένα δεδομένα από αυτά τα προφίλ χρησιμοποιούνται για τη δημιουργία συστάσεων. Επειδή τα προφίλ περιέχουν πολλά χαρακτηριστικά, πολλά από τα οποία έχουν κατακερματισμένα ή ελλιπή δεδομένα (Herlocker, J.L., Konstan, J. A. and Riedl, J. 2000), το να βρεθούν κατάλληλες ομοιότητες είναι συχνά δύσκολο. Για να ξεπεραστούν αυτά τα προβλήματα, τα σημερινά συστήματα (όπως το MovieLens) χρησιμοποιούν στοχαστικά και ευρετικά μοντέλα (heuristic models) για να επιταχύνουν και να βελτιώσουν την ποιότητα του ταιριάσματος των προφίλ. Το συγκεκριμένο σύστημα εφαρμόζει έναν αλγόριθμο βελτιστοποίησης σμήνους σωματιδίων στο πρόβλημα της αντιστοίχισης προφίλ.

Στην παρούσα έρευνα, χρησιμοποιήθηκε το σύνολο δεδομένων MovieLens για τα αρχικά πειράματα. Περιέχει λεπτομέρειες 943 χρηστών, καθεμία με πολλές παραμέτρους ή χαρακτηριστικά: δημογραφικά στοιχεία όπως η ηλικία, το φύλο και η επάγγελμα συλλέγονται όταν καταχωρείται ένας νέος χρήστης στο σύστημα. Κάθε φορά που υποβάλλεται μια ψηφοφορία από το χρήστη, καταγράφεται στη βάση δεδομένων με μια χρονική σήμανση. Οι πληροφορίες της ταινίας στο σύνολο των δεδομένων περιλαμβάνουν είδη και ημερομηνίες κυκλοφορίας στον κινηματογράφο και σε βίντεο. Και τα 2 συστήματα σύστασης PSO και GA χρησιμοποιούν 22 χαρακτηριστικά από αυτό το σύνολο δεδομένων: βαθμολογία ταινίας, ηλικία, φύλο, επάγγελμα και 18 είδη ταινιών: δράσης, περιπέτειας, animation, παιδική, κωμωδία, αστυνομική, ντοκιμαντέρ, δράμα, φαντασίας, film-noir, τρόμου, μουσικές, μυστηρίου, ρομαντική, επιστημονικής φαντασίας, θρίλερ, πολεμική, western.

A. προφίλ Generator

Πριν να μπορούν να γίνουν συστάσεις, τα δεδομένα της ταινίας πρέπει πρώτα να επεξεργαστούν σε ξεχωριστά προφίλ, ένα για κάθε άτομο, ορίζοντας τις προτιμήσεις αυτού του ατόμου σχετικά με τις ταινίες.

Ένα προφίλ για τον χρήστη j , ορίζεται σαν $profile(j)$, παρουσιάζεται σαν ένας πίνακας 22 τιμών για τα 22 προς εξέταση χαρακτηριστικά. Το προφίλ έχει δύο μέρη: ένα μεταβλητό μέρος (η βαθμολογία, η οποία αλλάζει σύμφωνα με την ταινία που εξετάζεται εκείνη τη στιγμή), και ένα σταθερό τμήμα (οι άλλες 21 τιμές, οι οποίες ανακτώνται μόνο μία φορά κατά την έναρξη του προγράμματος). Επειδή ο χρήστης j μπορεί να έχει βαθμολογήσει πολλές διαφορετικές ταινίες, ορίζουμε το $profile(j,i)$ να σημαίνει το προφίλ του χρήστη j για την ταινία i , σχήμα 1.

1 Rating	2 Age	3 Gender	4 Occupation	..22 18 Genre frequencies
5	23	0	45	000000100010000000

Σχήμα 1: προφίλ (j, i) - προφίλ για το χρήστη j που έχει βαθμολογήσει την ταινία i με 5.

Από τη στιγμή που δημιουργηθούν τα προφίλ, μπορεί να ξεκινήσει η διαδικασία σύστασης. Δεδομένου ενός ενεργού χρήστη A , πρέπει να βρεθούν ένα σύνολο προφίλ παρόμοιων με το $Profile(A)$.

4.3.1 Neighbourhood Selection

Η επιτυχία ενός συστήματος συνεργατικής διοίκησης είναι ανάλογη με την αποτελεσματικότητα του αλγορίθμου στην εύρεση του συνόλου ή της γειτονιάς των προφίλ που μοιάζουν περισσότερο με το προφίλ του ενεργού χρήστη. Είναι σημαντικό ότι, για μια συγκεκριμένη μέθοδο γειτονιάς, μόνο τα καλύτερα ή τα πιο κοντινά προφίλ επιλέγονται και χρησιμοποιούνται για να δημιουργηθούν νέες συστάσεις για τον χρήστη. Υπάρχει μικρή ανοχή για ανακρίβειες ή άσχετες προβλέψεις.

Ο αλγόριθμος επιλογής γειτονιάς αποτελείται από τρία κύρια μέρη: Επιλογή προφίλ (Profile Selection), ταιρίασμα προφίλ (Profile Matching) και συλλογή βέλτιστου Προφίλ (Best Profile Collection).

1) Επιλογή προφίλ

Σε έναν ιδανικό κόσμο, θα χρησιμοποιούνταν ολόκληρη η βάση δεδομένων των προφίλ για την επιλογή του καλύτερου δυνατού προφίλ. Ωστόσο, αυτό δεν είναι πάντα μια εφικτή επιλογή, ειδικά όταν το σύνολο των δεδομένων είναι πολύ μεγάλο ή εάν δεν υπάρχουν διαθέσιμοι πόροι. Ως αποτέλεσμα, τα περισσότερα συστήματα επιλέγουν τυχαία δειγματοληψία και η διαδικασία αυτή είναι ευθύνη του τμήματος επιλογής προφίλ του αλγορίθμου.

Αυτό το έργο διερευνά δύο μεθόδους για την επιλογή προφίλ:

1. Καθορισμένη (Fixed): οι πρώτοι n χρήστες από τη βάση δεδομένων χρησιμοποιούνται πάντα σε κάθε πείραμα
2. Τυχαία (Random): επιλέγονται τυχαία n χρήστες από τη βάση δεδομένων, όπου $n = 10$ ή 50 στα πειράματα μας

4.3.2 Ταιρίασμα προφίλ

Μετά την επιλογή του προφίλ, η διαδικασία ταιριάσματος προφίλ υπολογίζει στη συνέχεια την απόσταση ή την ομοιότητα μεταξύ των επιλεγμένων προφίλ και του ενεργού προφίλ του χρήστη χρησιμοποιώντας μια συνάρτηση απόστασης. Αυτή η έρευνα επικεντρώθηκε σε αυτό το έργο ταιριάσματος προφίλ,

δηλαδή, χρησιμοποίησε τον PSO αλγόριθμο για την βελτίωση του ταιριάσματος προφίλ για κάθε ενεργό χρήστη.

Από την ανάλυση των Breese et. al (1998), φαίνεται ότι τα περισσότερα τρέχοντα συστήματα σύστασης χρησιμοποιούν τυπικούς αλγόριθμους που λαμβάνουν υπόψη τους μόνο την "πληροφορία ψηφοφορίας" (voting information) σαν το χαρακτηριστικό βάσει του οποίου γίνεται η σύγκριση μεταξύ των δύο προφίλ. Ωστόσο στην πραγματική ζωή, ο τρόπος με τον οποίο λέγεται ότι δύο άτομα μοιάζουν, δεν βασίζεται αποκλειστικά και μόνο στο αν έχουν θετική γνώμη σχετικά με ένα συγκεκριμένο θέμα, π.χ., στην αξιολόγηση ταινιών, αλλά και σε άλλους παράγοντες, όπως το υπόβαθρό τους και τα προσωπικά τους στοιχεία.

Εάν αυτό εφαρμοστεί στο ταιρίασμα του προφίλ, πληροφορίες σχετικά με δημογραφικά δεδομένα και δεδομένα του τρόπου ζωής, τα οποία περιλαμβάνουν την ηλικία του χρήστη, το φύλο του και τις προτιμήσεις του σχετικά με τα είδη των ταινιών, πρέπει επίσης να λαμβάνονται υπόψη. Κάθε χρήστης δίνει μια διαφορετική σημασία ή προτεραιότητα σε κάθε χαρακτηριστικό γνώρισμα. Οι προτεραιότητες αυτές μπορούν να ποσοτικοποιηθούν ή να απαριθμηθούν. Αυτό γίνεται με τα βάρη των χαρακτηριστικών γνωρισμάτων. Για παράδειγμα, εάν ένας αρσενικός χρήστης προτιμά να του δοθούν συστάσεις με βάση τις γνώμες άλλων ανδρών, στη συνέχεια το βάρος για τα χαρακτηριστικά του φύλου θα είναι υψηλότερο από αυτό των άλλων χαρακτηριστικών. Προκειμένου να εφαρμόσουμε πραγματικά ένα σύστημα εξατομικευμένων συστάσεων, αυτά τα βάρη πρέπει να καταγραφούν, να βελτιστοποιηθούν και να καθοριστούν με ακρίβεια για να απεικονίσουν την προτίμηση κάθε χρήστη. Η προσέγγιση αυτή δείχνει πώς μπορούν να βρεθούν τέτοια βάρη χρησιμοποιώντας την τεχνική βελτιστοποίησης σωματιδίων.

Μια πιθανή λύση στο πρόβλημα της βελτιστοποίησης των βαρών των χαρακτηριστικών $w(A)$ για τον ενεργό χρήστη, όπου το A παρουσιάζεται ως ένα σύνολο βαρών, όπως φαίνεται παρακάτω στο σχήμα 2.

w_1	w_2	w_3	...	w_{22}
-------	-------	-------	-----	----------

Σχήμα 2: Το σύνολο των βαρών συντελεστών στάθμισης που αντιπροσωπεύει την προτίμηση του χρήστη.

όπου w_f είναι το βάρος που σχετίζεται με το χαρακτηριστικό γνώρισμα f . Κάθε σύνολο περιέχει 22 τιμές στάθμισης, που αντιπροσωπεύουν μια θέση ενός σωματιδίου στο 22-διάστατων διάστημα αναζήτησης. Καθώς τα σωματίδια κινούνται γύρω στο χώρο, αυτές οι τιμές προσαρμόζονται συνεχώς (περιγράφεται στο τμήμα Δ) προκειμένου να βρεθεί το καλύτερο σωματίδιο με ένα σύνολο συντελεστών στάθμισης που να περιγράφουν με ακρίβεια την προτίμηση του ενεργού χρήστη.

Στη συνέχεια μπορεί να γίνει η σύγκριση μεταξύ των δύο προφίλ χρησιμοποιώντας μια τροποποιημένη συνάρτηση Ευκλείδειας απόστασης, που θα λαμβάνει υπόψη της πολλά χαρακτηριστικά. Το μέγεθος $\text{Euclidean}(A, j)$ είναι η ομοιότητα μεταξύ ενεργού χρήστη A και χρήστη j :

$$\text{euclidean}(A, j) = \sqrt{\sum_{i=1}^z \sum_{f=1}^{22} w_f * \text{diff}_{i,f}(A, j)^2}$$

Όπου: A είναι ο ενεργός χρήστης

j είναι ένας χρήστης που παρέχεται από τη διαδικασία επιλογής προφίλ, όπου $j \neq A$

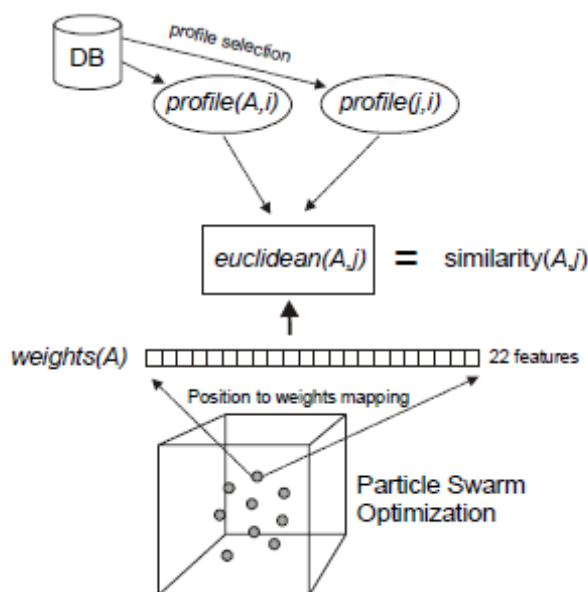
z είναι το πλήθος των κοινών ταινιών που αξιολόγησαν οι χρήστες A και j

w_f είναι το βάρος του ενεργού χρήστη για το χαρακτηριστικό f

i είναι μια κοινή ταινία, όπου υπάρχουν τα $profile(A,i)$ και $profile(j,i)$

$diff_i,f(A,j)$ είναι η διαφορά στην τιμή του προφίλ για το χαρακτηριστικό f ανάμεσα στους χρήστες A και j για την ταινία i

Πριν γίνουν όλοι αυτοί οι υπολογισμοί, οι τιμές των προφίλ κανονικοποιούνται για να εξασφαλιστεί ότι κυμαίνονται ανάμεσα στο 0 και στο 1. Όταν ο συντελεστής βαρύτητας για κάποιο χαρακτηριστικό είναι 0, αυτό το χαρακτηριστικό αγνοείται. Με τον τρόπο αυτό, προσαρμόζεται η επιλογή χαρακτηριστικού στις προτιμήσεις κάθε χρήστη. Η διαφορά στις τιμές των προφίλ για το επάγγελμα είναι 0, αν οι δύο χρήστες έχουν το ίδιο επάγγελμα ή 1 σε διαφορετική περίπτωση.



Σχήμα 3: Υπολογισμός της ομοιότητας μεταξύ A και j

4.3.3 Best Profile Collection

Αφού βρεθούν οι Ευκλείδειες αποστάσεις $euclidean(A, j)$ μεταξύ των $profile(A)$ και $profile(j)$ για όλες τις τιμές j μέσω της διαδικασίας επιλογής προφίλ, καλείται ο αλγόριθμος "best profile collection". Αυτός κατατάσσει κάθε $profile(j)$ σύμφωνα με την ομοιότητά του με το $profile(A)$. Στη συνέχεια, το σύστημα απλά επιλέγει τους χρήστες των οποίων η Ευκλείδεια απόσταση είναι πάνω από μια συγκεκριμένη τιμή κατωφλίου (που θεωρείται πιο όμοια με τον ενεργό χρήστη) σαν τη γειτονιά του A . Αυτή η τιμή είναι μια σταθερά του συστήματος που μπορεί να αλλάξει.

4.3.4 Κάνοντας μια σύσταση

Για να γίνει μια σύσταση, δεδομένου ενός ενεργού χρήστη A και μιας γειτονιάς αποτελούμενης από παρόμοια προφίλ με αυτό του A , είναι απαραίτητο να βρεθούν ταινίες που να τις έχουν δει οι χρήστες στη γειτονιά (και να τους έχουν αρέσει) και τις οποίες ο ενεργός χρήστης δεν έχει δει. Στη συνέχεια, αυτές παρουσιάζονται στον ενεργό χρήστη μέσω μιας διαπαφής χρήστη. Επειδή η γειτονιά περιέχει τους

χρήστες που είναι πιο παρόμοιοι με τον A (χρησιμοποιώντας, στην περίπτωση μας, τις συγκεκριμένες προτιμήσεις του A μέσω των επιτευχθέντων τιμών στάθμισης), ταινίες που άρεσαν σ' αυτούς τους χρήστες, έχουν μια εύλογη πιθανότητα να αρέσουν στον A.

Δ. PSO αλγόριθμος

Ο αλγόριθμος PSO έχει χρησιμοποιηθεί για την απόκτηση των βαρών των χαρακτηριστικών του ενεργού χρήστη, και ως εκ τούτου, βοηθά στην προσαρμογή της συνάρτησης ταιριάσματος στη συγκεκριμένη προσωπικότητα και στις προτιμήσεις του χρήστη.

1) Δυναμική των Σωματιδίων (Particle Dynamics)

Επιλέχθηκε ένας συμβατικός PSO αλγόριθμος (Eberhart, R.C. and Shi, Y. 2001), ο οποίος έχει έναν χώρο 22 διαστάσεων, ο κάθε άξονας έχει τιμές που κυμαίνονται από 0 έως 255 (που αντιστοιχεί στην απλή μη προσημασμένη δυαδική γενετική κωδικοποίηση με 8 bits για κάθε ένα από τα 22 γονίδια, που χρησιμοποιούνται για την εφαρμογή του γενετικού αλγόριθμου (GA) recommender συστήματος. Όπως προαναφέρθηκε, κάθε σωματίδιο έχει μια θέση και μια ταχύτητα (οι τιμές τους αρχικά είναι τυχαίες). Η θέση με την υψηλότερη βαθμολογία καταλληλότητας σε κάθε επανάληψη έχει οριστεί να είναι για το σύνολο του σμήνους η global best (gbest) θέση, προς την οποία μετακινούνται τα άλλα σωματίδια. Επιπλέον, κάθε σωματίδιο διατηρεί την καλύτερη θέση που έχει επισκεφθεί, γνωστή και ως personal best (pbest) του σωματιδίου. Η δυναμική των σωματιδίων διέπεται από τους ακόλουθους κανόνες, οι οποίοι ενημερώνουν τις θέσεις και τις ταχύτητες των σωματιδίων:

$$v_i = wv_i + c_1r_1(x_{pbest,i} - x_i) + c_2r_2(x_{gbest} - x_i)$$

$$if(|v_i| > v_{max}) \quad v_i = (v_{max} / |v_i|)v_i$$

$$x_i = x_i + v_i$$

Όπου

x_i είναι η τρέχουσα θέση του σωματιδίου i

x_{pbest} είναι η καλύτερη θέση που αποκτήθηκε από το σωματίδιο i

x_{gbest} είναι η καλύτερη συνολική θέση του σμήνους

v_i είναι η ταχύτητα του σωματιδίου i

w είναι το τυχαίο βάρος της αδράνειας ανάμεσα σε 0.5 και 1

c_1 και c_2 είναι σταθερές με τιμή 1.494

r_1 και r_2 είναι τυχαίοι αριθμοί ανάμεσα στο 0 και το 1

2) Η συνάρτηση καταλληλότητας (Fitness Function)

Ο υπολογισμός της καταλληλότητας για αυτήν την εφαρμογή δεν είναι συνηθισμένος. Ένα σύνολο των συντελεστών στάθμισης μπορεί να υπολογιστεί από τις τιμές της θέσης. Πρώτον, η σημασία των 18 ειδών μειώνεται κατά ένα δεδομένο παράγοντα, το μέγεθος της μείωσης του βάρους. Αυτό γίνεται επειδή τα 18 είδη μπορεί να θεωρηθούν διαφορετικές κατηγορίες ενός ενιαίου μεγαλύτερου χαρακτηριστικού, του Είδους. Η μείωση του αποτελέσματος αυτών των βαρών προορίζεται ως εκ τούτου, να δώσει στα άλλα μη συνδεδεμένα χαρακτηριστικά (βαθμολογία ταινίας, ηλικία, φύλο, επάγγελμα) μια πιο ίση ευκαιρία να χρησιμοποιηθούν. Δεύτερον, η συνολική αξία της θέσης υπολογίζεται στη συνέχεια αθροίζοντας τις τιμές της θέσης σε όλους τους 22 άξονες. Τέλος, η σταθμισμένη τιμή για κάθε χαρακτηριστικό μπορεί να βρεθεί διαιρώντας την πραγματική τιμή με τη συνολική τιμή. Το άθροισμα όλων των βαρών στη συνέχεια είναι ίσο με 1.

Η τρέχουσα θέση κάθε σωματιδίου (σύνολο των συντελεστών στάθμισης) στο σμήνος πρέπει να χρησιμοποιηθεί από τη διαδικασία ταιριάσματος προφίλ στο πλαίσιο του συστήματος recommender. Έτσι, το σύστημα recommender θα πρέπει να ξαναεκτελεστεί για το σύνολο των δεδομένων MovieLens για κάθε νέα θέση, για τον υπολογισμό της καταλληλότητάς της.

Αποφασίστηκε η επαναδιατύπωση του προβλήματος σαν μια εργασία επιτηρούμενης μάθησης. Όπως περιγράφηκε προηγουμένως, δεδομένου του ενεργού χρήστη A και ενός συνόλου γειτονικών

προφίλ, μπορούν να γίνουν συστάσεις για τον A. Επιπρόσθετα με τις συστάσεις αυτές, είναι δυνατόν να προβλέψουμε τι μπορεί να σκεφτεί ο A για αυτές. Για παράδειγμα, εάν μια ορισμένη ταινία προτείνεται επειδή την είδαν παρόμοιοι χρήστες, αλλά αυτοί οι χρήστες την αξιολόγησαν σαν "Μέτρια", είναι πιθανό ότι ο ενεργός χρήστης μπορεί επίσης να αξιολογήσει την ταινία σαν "Μέτρια". Ως εκ τούτου, για το σύνολο δεδομένων MovieLens, ήταν δυνατό για το σύστημα και να συστήνει νέες ταινίες και να προβλέπει πώς θα αξιολογήσει ο ενεργός χρήστης κάθε ταινία, εάν πάει να τη δει.

Ο υπολογισμός της προβλεφθείσας ψηφοφορίας που χρησιμοποιήθηκε σε αυτό το έγγραφο έχει παρθεί από (Breese, J.S., Heckerman, D. and Kadie, C. 1998.) και τροποποιήθηκε έτσι ώστε η συνάρτηση Ευκλείδειας απόστασης (Profile Matching) τώρα αντικαθιστά το βάρος στην αρχική εξίσωση. Η προβλεπόμενη ψηφοφορία, $predict_vote(A,i)$, για το A στο στοιχείο i, μπορεί να οριστεί ως:

$$predict_vote(A,i) = mean_A + k \sum_{j=1}^n euclidean(A,j)(vote(j,i) - mean_j)$$

όπου: $mean_j$ είναι η μέση ψηφοφορία για το χρήστη j

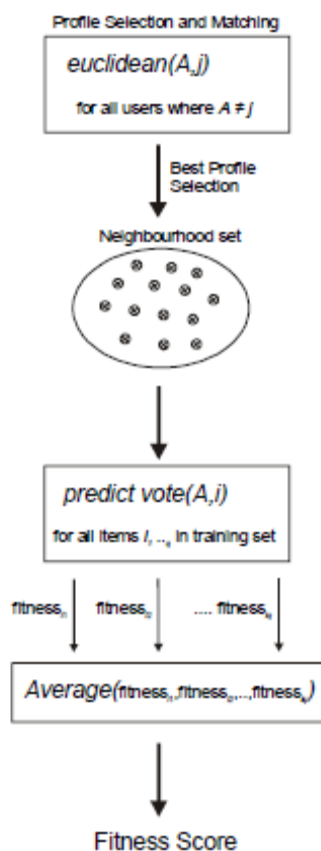
k είναι ένας παράγοντας κανονικοποίησης ώστε το άθροισμα της ευκλείδειας απόστασης να είναι ίσο με 1.

$vote(j,i)$ είναι η πραγματική ψηφοφορία που ο χρήστης j έδωσε για το στοιχείο i

n είναι το μέγεθος της γειτονιάς

Όλα τα είδη ταινιών που έχει δει ο ενεργός χρήστης διαμερίζονται τυχαία σε δύο σύνολα δεδομένων: ένα σύνολο εκπαίδευσης (1/3) και ένα σύνολο δοκιμής (2/3). Για να υπολογιστεί ένα μέτρο καταλληλότητας για την τρέχουσα θέση ενός σωματιδίου, το σύστημα recommender βρίσκει ένα σύνολο προφίλ γειτονιάς για τον ενεργό χρήστη, όπως περιγράφεται στην ενότητα επιλογή γειτονιάς.

Οι εκτιμήσεις των χρηστών στο σύνολο της γειτονιάς χρησιμοποιούνται για να υπολογιστεί η προβλεφθείσα εκτίμηση για τον ενεργό χρήστη για κάθε ταινία στο σύνολο της εκπαίδευσης. Επειδή ο ενεργός χρήστης έχει ήδη αξιολογήσει τις ταινίες, είναι δυνατό να συγκριθεί η πραγματική εκτίμηση με την προβλεφθείσα εκτίμηση. Έτσι, ο μέσος όρος των διαφορών μεταξύ της πραγματικής και της προβλεπόμενης ψήφου για όλα τα είδη στο σύνολο εκπαίδευσης χρησιμοποιείται σαν βαθμολογία καταλληλότητας για την καθοδήγηση της μελλοντικής διαδικασίας σχηματισμού σμήνους, βλέπε σχήμα 4.



Σχήμα 4: Εύρεση της τιμής καταλληλότητας της τρέχουσας θέσης ενός σωματιδίου (βάρη των χαρακτηριστικών του ενεργού χρήστη)

4.3.5 Πειράματα

Τέσσερα σύνολα πειραμάτων σχεδιάστηκαν για την παρακολούθηση της διαφοράς της απόδοσης μεταξύ PSO, GA (Ujjin, S. and Bentley, P. J. 2002) και ενός τυπικού, μη προσαρμόσιμου recommender συστήματος με βάση τον Pearson αλγόριθμο (Breese, J.S., Heckerman, D. and Kadie, C. 1998). Σε κάθε σύνολο πειραμάτων, οι προβλεπόμενες ψηφοφορίες όλων των ταινιών στο σύνολο δοκιμής (τα στοιχεία που ο ενεργός χρήστης έχει βαθμολογήσει, αλλά δεν χρησιμοποιήθηκαν στον υπολογισμό της καταλληλότητας), υπολογίστηκαν χρησιμοποιώντας τα τελικά βάρη χαρακτηριστικών για αυτή την εκτέλεση. Στη συνέχεια, αυτές οι ψηφοφορίες συγκρίθηκαν με αυτές που παράγονται από τον απλό Pearson αλγόριθμο και το σύστημα γενετικού αλγόριθμου (GA σύστημα).

Τα τέσσερα πειράματα υπολόγισαν δύο μεταβλητές συστήματος για την αξιολόγηση των επιπτώσεών τους στις επιδόσεις του συστήματος: την επιλογή προφίλ (ο τρόπος με τον οποίο επιλέγονται τα προφίλ από τη βάση δεδομένων), και το μέγεθος της γειτονίας.

Τα τέσσερα σύνολα των πειραμάτων έχουν ως εξής:

Πείραμα 1: Καθένας από τους πρώτους 10 χρήστες επιλέχθηκε με τη σειρά ως ενεργός χρήστης, και οι πρώτοι 10 χρήστες (σταθερό) χρησιμοποιήθηκαν για την παροχή συστάσεων.

Πείραμα 2: Καθένας από τους πρώτους 50 χρήστες επιλέχθηκε με τη σειρά ως ενεργός χρήστης, και οι πρώτοι 50 χρήστες (σταθερό) χρησιμοποιήθηκαν για την παροχή συστάσεων.

Πείραμα 3: Καθένας από τους πρώτους 10 χρήστες επιλέχθηκε με τη σειρά ως ενεργός χρήστης, και 10 από τους 944 χρήστες επιλέχθηκαν τυχαία και χρησιμοποιήθηκαν για την παροχή συστάσεων (οι ίδιοι 10 χρησιμοποιήθηκαν για κάθε εκτέλεση και για τα τρία συστήματα).

Πείραμα 4: Καθένας από τους πρώτους 50 χρήστες επιλέχθηκε με τη σειρά ως ενεργός χρήστης, και 50 από τους 944 χρήστες επιλέχθηκαν τυχαία και χρησιμοποιήθηκαν για την παροχή συστάσεων (οι ίδιοι 50 χρησιμοποιήθηκαν για κάθε εκτέλεση και για τα τρία συστήματα).

Αυτά έγιναν σε δύο εκδόσεις του συστήματος:

Zero tolerance – η ακρίβεια του συστήματος διαπιστώνεται από τον υπολογισμό του ποσοστού του αριθμού των αξιολογήσεων που το σύστημα πρόβλεψε σωστά μέσα από το συνολικό αριθμό των διαθέσιμων αξιολογήσεων από τον τρέχοντα ενεργό χρήστη. Τα αποτελέσματα για πειράματα 1 έως 4 με μηδενική ανοχή εμφανίζονται στα σχήματα 5 έως 8 αντίστοιχα.

At-Most-One tolerance – ίδιο με το μηδενικής ανοχής σύστημα, αλλά αν η διαφορά μεταξύ των προβλεπόμενων και πραγματικών αξιολογήσεων είναι μικρότερη από ή ίση με 1, τότε αυτή η προβλεπόμενη αξιολόγηση θεωρείται ότι είναι σωστή. Αυτό λαμβάνει υπόψη την ευπάθεια του συστήματος αξιολόγησης – οι αξιολογήσεις ενός στοιχείου μπορεί να διαφέρουν εξαρτώμενες από πολλούς παράγοντες, όπως η ώρα της ημέρας και η διάθεση του χρήστη τη δεδομένη χρονική στιγμή. Τα σχήματα 9 έως 12 εμφανίζουν τα αποτελέσματα για τα πειράματα 1 έως 4 με ανοχή at-most-one.

Σε όλα τα πειράματα διατηρήθηκαν ίδιες οι παρακάτω τιμές παραμέτρων:

1. μέγεθος σμήνους/πληθυσμού = 75. Ο αριθμός των σωματιδίων/ατόμων στο σμήνος/πληθυσμό σε κάθε γενιά.
2. Μέγιστος αριθμός επαναλήψεων για κάθε εκτέλεση = 300. Εάν ο αριθμός των επαναλήψεων φτάσει αυτή την τιμή και δεν έχει βρεθεί η λύση, η καλύτερη λύση γι' αυτή την επανάληψη χρησιμοποιείται ως το τελικό αποτέλεσμα.
3. μέγεθος ελάττωσης βάρους = 4. Ο συντελεστής κλίμακας για τα 18 είδη.
4. αριθμός των εκτελέσεων = 30. Ο αριθμός των φορών που το σύστημα έτρεξε για κάθε ενεργό χρήστη.

Ο αλγόριθμος Pearson που χρησιμοποιήθηκε στα πειράματα βασίζεται στον αλγόριθμο k Nearest Neighbour. Ένας συντελεστής συσχέτισης, όπως φαίνεται παρακάτω, χρησιμοποιείται σαν η αντίστοιχη συνάρτηση για την επιλογή των k χρηστών που μοιάζουν περισσότερο με τον ενεργό χρήστη για να δώσουν προβλέψεις. Αυτός αντικαθιστά την Ευκλείδεια συνάρτηση που περιγράφηκε παραπάνω, όλες οι άλλες λεπτομέρειες παραμένουν ίδιες.

$$\text{correlation}(A, j) = \frac{\sum_{i=1}^k (\text{vote}(A, i) - \text{mean}_A)(\text{vote}(j, i) - \text{mean}_j)}{\sqrt{\sum_{i=1}^k (\text{vote}(A, i) - \text{mean}_A)^2 (\text{vote}(j, i) - \text{mean}_j)^2}}$$

Το Σύστημα σύστασης γενετικού αλγόριθμου χρησιμοποιεί ένα ελιτίστικο γενετικό αλγόριθμο, όπου το ένα τέταρτο των καλύτερων ατόμων στον πληθυσμό διατηρήθηκαν για την επόμενη γενιά. Κατά τη δημιουργία μιας νέας γενιάς, άτομα επιλέγονται τυχαία από το καλύτερο 40% του συνόλου του πληθυσμού προκειμένου να γίνουν γονείς. Δύο απόγονοι παράγονται από κάθε ζεύγος γονέων, χρησιμοποιώντας διασταύρωση single-point με πιθανότητα 1. Εφαρμόστηκε μετάλλαξη στον γονότυπο με πιθανότητα 0,01. Μια απλή μη προσημασμένη δυαδική γενετική κωδικοποίηση χρησιμοποιήθηκε κατά την εφαρμογή, χρησιμοποιώντας 8 bits για κάθε ένα από τα 22 γονίδια που αντιπροσώπευαν τα 22 υπό θεώρηση χαρακτηριστικά.

A. Αποτελέσματα

Ενώ οι προβλέψεις που υπολογίζονται με τον αλγόριθμο Pearson παραμένουν πάντα οι ίδιες με δεδομένες τις ίδιες τιμές των παραμέτρων, αυτές που προκύπτουν από τους αλγόριθμους PSO και GA ποικίλουν ανάλογα με τα βάρη των χαρακτηριστικών μιας εκτέλεσης. Από τις 30 εκτελέσεις για κάθε ενεργό χρήστη σε κάθε πείραμα, η εκτέλεση με τα καλύτερα βάρη χαρακτηριστικών (που έδωσαν το μεγαλύτερο ποσοστό σωστών προβλέψεων) επιλέχθηκε και σχεδιάστηκε σε αντιδιαστολή με τον αλγόριθμο Pearson.

Εξετάζοντας τα αποτελέσματα που προκύπτουν από τα πειράματα, έγιναν οι ακόλουθες παρατηρήσεις:

- Γενικά ο PSO αλγόριθμος απέδωσε σημαντικά καλά σε σύγκριση με τα άλλα δύο συστήματα, σχήματα 5 έως 8 για αποτελέσματα από τα πειράματα 1 έως 4.
- Όταν χρησιμοποιήθηκε η τυχαία επιλογή προφίλ, στις περισσότερες περιπτώσεις η ακρίβεια της πρόβλεψης για τον PSO αυξήθηκε σημαντικά και ξεπέρασε τα άλλα δύο συστήματα. Τα σχήματα 7 και 8 δείχνουν τα αποτελέσματα των πειραμάτων 3 και 4.
- Η απόδοση του συστήματος PSO βελτιώθηκε σε μεγάλο βαθμό όντας 40-50% ακριβής με μηδενικό επίπεδο ανοχής και 60-100% ακριβής με χρήση του at-most-one tolerance level. Το Σχήμα 13 δείχνει την ακρίβεια μέσου όρου πρόβλεψης και για τα τέσσερα πειράματα και για τα δύο επίπεδα ανοχής.
- Η ταχύτητα στην εκτέλεση του PSO συστήματος ήταν 10% μεγαλύτερη από αυτή του GA.

B. Ανάλυση των αποτελεσμάτων

Διαπιστώθηκε ότι ο PSO αλγόριθμος εκτελείται πολύ καλά σε σύγκριση με τα άλλα δύο συστήματα για τα τέσσερα πειράματα και στα δύο επίπεδα ανοχής. Ωστόσο, όταν ο αριθμός των χρηστών αυξάνει, η πιθανότητα εύρεσης ενός καλύτερου ταιριάσματος προφίλ θα έπρεπε να είναι υψηλότερη και ως εκ τούτου η ακρίβεια των προβλέψεων θα έπρεπε να αυξηθεί, αυτό εξακολουθεί να ισχύει για το GA σύστημα, αλλά δεν ισχύει για τον PSO recommender. Μία εξήγηση για αυτή τη μείωση του επιπέδου μέσης ακρίβειας με την αύξηση του αριθμού των χρηστών είναι ότι, αρχικά οι χρήστες που έχουν επιλεγεί για να συμπεριληφθούν στη γειτονιά για να δώσουν συστάσεις, είναι πολύ παρόμοιοι με τον ενεργό χρήστη. Καθώς ο αριθμός των χρηστών αυξάνει, εξετάζονται περισσότεροι χρήστες και αυτό μερικές φορές θα μπορούσε να οδηγήσει σε λιγότερο παρόμοιους χρήστες που προστίθενται στη γειτονιά και ως εκ τούτου, να μειώσει συνολικά την ακρίβεια της πρόβλεψης.

Αποτελέσματα για το PSO από τα πειράματα, 3 και 4 επιβεβαιώνουν αυτό που ισχύει και για τον GA recommender (Ujjin, S. and Bentley, P. J. 2002.), ότι η τυχαία δειγματοληψία είναι καλύτερη για την επιλογή των χρηστών. Αυτό οφείλεται στο ότι επιτρέπει την αναζήτηση για μια μεγαλύτερη ποικιλία προφίλ. Επιπλέον, εξετάστηκαν τα μέσα και χειρότερα αποτελέσματα και συγκρίθηκαν μεταξύ των τριών μεθόδων. Διαπιστώθηκε ότι ο PSO πετυχαίνει τις καλύτερες επιδόσεις στα περισσότερα πειράματα σε σύγκριση με τα συστήματα GA και PA. Η μοναδική περίπτωση κατά την οποία ο PSO δεν πέτυχε την υψηλότερη απόδοση ήταν στο πείραμα 2 όπου έδωσε χειρότερα αποτελέσματα, τονίζοντας το πρόβλημα της αύξησης του αριθμού των χρηστών που αναφέρθηκε παραπάνω. Ωστόσο, αυτό το πρόβλημα δεν έχει καμία επίπτωση στις επιδόσεις του PSO στο πείραμα 4 όταν χρησιμοποιήθηκε τυχαία δειγματοληψία.

Όπως αναφέρθηκε προηγουμένως, μόνο η/οι εκτέλεση/εκτελέσεις με τα καλύτερα βάρη χαρακτηριστικών για κάθε ενεργό χρήστη λήφθηκε υπόψη για αυτή την ανάλυση. Στον GA recommender όταν περισσότερες από μια εκτελέσεις για έναν ενεργό χρήστη δίνουν την ίδια καλύτερη επίδοση (ο μέγιστος αριθμός ψήφων που έχουν προβλεφθεί σωστά), τα αποτελέσματα δείχνουν ότι έχει επιλεγεί το ίδιο σύνολο χρηστών για τη γειτονιά για να δώσουν συστάσεις (Ujjin, S. and Bentley, P. J. 2002). Ωστόσο, αυτό δεν ισχύει για το σύστημα PSO: πολλές εκτελέσεις που πέτυχαν τις ίδιες καλύτερες επιδόσεις, δεν επέλεξαν το ίδιο σύνολο χρηστών. Εξετάζοντας το ιστορικό των διαδρομών των σωματιδίων, όταν περισσότερα από 1 σωματίδια επιτυγχάνουν την ίδια καλύτερη επίδοση, η global best θέση επιλέγεται τυχαία από ένα από αυτά τα σωματίδια και είναι αυτή η θέση προς την οποία κινούνται τα άλλα σωματίδια. Είναι δυνατόν τα άλλα "καλύτερα" σωματίδια που δεν επιλέχθηκαν να έδιναν καλύτερη λύση, εάν το σμήνος είχε την ευκαιρία να διερευνήσει τον περιβάλλοντα χώρο τους.

4.3.6 Συμπεράσματα

Αυτή η εργασία έχει δείξει πώς η βελτιστοποίηση σμήνους σωματιδίων θα μπορούσε να χρησιμοποιηθεί για την τελειοποίηση ενός αλγόριθμου ταιριάσματος προφίλ σε ένα σύστημα recommender, προσαρμόζοντάς τον στις προτιμήσεις των μεμονωμένων χρηστών. Πειράματα έδειξαν ότι το PSO σύστημα έδειξε μεγαλύτερη ακρίβεια πρόβλεψης από το σύστημα γενετικών αλγορίθμων και τον αλγόριθμο Pearson στις περισσότερες περιπτώσεις. Επιπλέον, σε σύγκριση με την προσέγγιση GA, ο αλγόριθμος PSO βρήκε την τελική λύση σημαντικά πιο γρήγορα, καθιστώντας τον έναν πιο αποτελεσματικό τρόπο για τη βελτίωση της απόδοσης, όπου η ταχύτητα των υπολογισμών παίζει ένα σημαντικό ρόλο στα συστήματα recommender.

ΚΕΦΑΛΑΙΟ 5

5. Συμπεράσματα

Στην παρούσα εργασία έγινε μια εκτενής παρουσίαση των Συστημάτων Σύστασης υλοποιημένων με αλγόριθμους βασισμένους στη Νοημοσύνη Σμήνους.

Τα συστήματα Recommender αναπτύχθηκαν σε απάντηση μιας προφανούς ανάγκης: το να βοηθηθούν οι άνθρωποι να αντιμετωπίσουν τον κόσμο της αφθονίας και της υπερφόρτωσης των πληροφοριών. Υπάρχει ένα σύνολο από τέσσερα σημαντικά ζητήματα για τα Συστήματα Σύστασης:

- ο τρόπος με τον οποίο λαμβάνονται και χρησιμοποιούνται τα δεδομένα προτίμησης,
- ο ρόλος που παίζουν οι άνθρωποι και οι υπολογισμοί, καθώς και οι σχετικοί τύποι επικοινωνίας,
- οι αλγόριθμοι για τη σύνδεση των ατόμων και τον υπολογισμό των συστάσεων και
- η παρουσίαση των συστάσεων στους χρήστες.

Τα Συστήματα σύστασης είναι νέου τύπου εργαλεία λογισμικού που βασίζονται στο Διαδίκτυο, σχεδιασμένα να βοηθούν τους χρήστες να βρουν το δρόμο τους μέσα στα σημερινά περίπλοκα online καταστήματα και δικτυακούς ιστότοπους ψυχαγωγίας. Η παρούσα εργασία περιγράφει Συστήματα Σύστασης, τα οποία χρησιμοποιούν τους αλγόριθμους βελτιστοποίησης σμήνους σωματιδίων (particle swarm optimization algorithm (PSO)) και ant colony optimization για να μάθουν τις προσωπικές προτιμήσεις των χρηστών και να παράσχουν προσαρμοσμένες συστάσεις στα μέτρα του καθενός.

Οι τρέχουσες προκλήσεις για τα συστήματα recommender αφορούν:

- 1) τη συγκέντρωση ατόμων σε κοινότητες ενδιαφέροντος. Μια σημαντική ανησυχία στην περίπτωση αυτή είναι ο σεβασμός της ιδιωτικής ζωής των ανθρώπων αυτών και
- 2) τη δημιουργία αλγορίθμων συστάσεων που συνδυάζουν πολλούς τύπους πληροφοριών, οι οποίοι έχουν αποκτηθεί κατά πάσα πιθανότητα από διαφορετικές πηγές σε διαφορετικές χρονικές στιγμές.

Κάθε μια προσέγγιση Recommender συστήματος έχει ορισμένες εγγενείς αδυναμίες. Συγκεκριμένα:

- Τα Content-based filtering Συστήματα σύστασης βασίζονται σε αντικειμενικές πληροφορίες για τα αντικείμενα. Αυτές οι πληροφορίες είτε εξάγονται αυτόματα από τις διάφορες πηγές (π.χ., ιστοσελίδες), είτε εισάγονται με το χέρι (π.χ., βάση δεδομένων προϊόντων). Εντούτοις, η επιλογή κάποιου στοιχείου είναι βασισμένη συνήθως σε υποκειμενικές ιδιότητες του στοιχείου (π.χ., ένα καλογραμμένο έγγραφο ή ένα προϊόν με μια πικάντικη γεύση). Επομένως, αυτές οι ιδιότητες, οι οποίες επηρεάζουν περισσότερο την επιλογή του χρήστη, δεν λαμβάνονται υπόψη.
- Ένα άλλο πρόβλημα, που έχει μελετηθεί εκτενώς, είναι της υπερεξειδίκευσης (overspecialization). Αυτά τα συστήματα δεν έχουν καμιά σύμφυτη μέθοδο για καλά ευρήματα από τυχαίες ανακαλύψεις (serendipity). Το σύστημα συστήνει περισσότερο αυτό που ο χρήστης έχει ήδη δει και έχει δείξει κάποια προτίμηση. Όταν το σύστημα μπορεί μόνο να συστήσει τα στοιχεία που σημειώνουν ιδιαίτερα υψηλή βαθμολογία σε σχέση με το προφίλ ενός χρήστη, ο χρήστης είναι περιορισμένος να δει στοιχεία παρόμοια με εκείνα που έχουν ήδη εκτιμηθεί.
- Οι εκτιμήσεις ενός χρήστη είναι ο μόνος παράγοντας που επηρεάζει τη μελλοντική απόδοση. Εντούτοις, μόνο μερικές εκτιμήσεις είναι διαθέσιμες λόγω της απροθυμίας των χρηστών να εκτελέσουν ενέργειες από τις οποίες δεν έχουν άμεσα οφέλη, καθώς και λόγω της χαμηλής αλληλεπίδρασης του χρήστη με το σύστημα. Επομένως, η ποιότητα της σύστασης δεν είναι πολύ ακριβής. Εντούτοις, αυτές οι ανεπάρκειες μπορούν να λυθούν με το συνδυασμό αυτής της προσέγγισης με collaborative filtering συστήματα.

Μειονεκτήματα του collaborative filtering συστήματος είναι τα εξής:

- Το early-rater πρόβλημα: όταν ένα νέο στοιχείο εμφανίζεται στη βάση δεδομένων δεν υπάρχει κανένας τρόπος να συστηθεί σε έναν χρήστη έως ότου αποκτηθούν περισσότερες πληροφορίες μέσω μιας άλλης εκτίμησης χρηστών, είτε βαθμολογώντας το, είτε καθορίζοντας με ποιά άλλα στοιχεία αυτό είναι παρόμοιο.
- Το πρόβλημα της σποραδικότητας (sparsity): ο στόχος των collaborative filtering συστημάτων είναι να ενισχυθούν οι άνθρωποι στο να εστιάσουν το ενδιαφέρον τους στην ανάγνωση εγγράφων (ή στην κατανάλωση στοιχείων). Έτσι, εάν ο αριθμός των χρηστών είναι μικρός σχετικά με τον όγκο των πληροφοριών στο σύστημα (επειδή υπάρχει μια πολύ μεγάλη ή μεταβαλλόμενη με γρήγορους ρυθμούς βάση δεδομένων), υπάρχει ένας κίνδυνος από τις εκτιμήσεις που γίνονται πάρα πολύ λίγες, πράγμα που κάνει ισχύη τη συλλογή των αξιούσστατων στοιχείων. Επίσης, η σποραδικότητα θέτει μια πραγματική υπολογιστική πρόκληση καθώς γίνεται δυσκολότερο να βρεθούν γείτονες και να δοθούν συστάσεις για τα στοιχεία, δεδομένου ότι πολύ λίγοι άνθρωποι έχουν κάνει αξιολογήσεις.
- Ένα άλλο πρόβλημα λογικής είναι ότι όταν οι προτιμήσεις ενός χρήστη ποικίλλουν από τον κανόνα, δεν υπάρχουν άλλοι χρήστες που να μοιράζονται τα ίδια γούστα, πράγμα που οδηγεί σε φτωχές συστάσεις.
- Η δυσκολία επίτευξης κρίσιμης μάζας συμμετεχόντων κάνει τα πειράματα collaborative filtering ακριβά. Η κρίσιμη εξάρτηση από το μέγεθος και τη σύνθεση του πληθυσμού των χρηστών επηρεάζει επίσης την ομάδα κοντινότερων γειτόνων του χρήστη. Σε μια κατάσταση στην οποία η ανατροφοδότηση αποτυγχάνει να αναγκάσει αυτήν την ομάδα των κοντινότερων γειτόνων να αλλάξει, η έκφραση της απέχθειας για ένα στοιχείο δεν θα αποτρέψει απαραίτητως το χρήστη από να λάβει παρόμοια στοιχεία στο μέλλον. Επιπλέον, η έλλειψη πρόσβασης στο περιεχόμενο των στοιχείων αποτρέπει τους παρόμοιους χρήστες από το να ταιριαστούν, εκτός και αν έχουν εκτιμήσει ακριβώς τα ίδια στοιχεία.

Τα υβριδικά συστήματα εκμεταλλεύονται τα χαρακτηριστικά γνωρίσματα των δύο αυτών συστημάτων. Αφ' ενός, τα καθαρά collaborative filtering συστήματα λύνουν τις ανεπάρκειες των αμιγώς content-based συστημάτων. Η πρώτη ανεπάρκεια των content-based συστημάτων είναι η έλλειψη υποκειμενικών στοιχείων για τα στοιχεία. Σε ένα collaborative filtering σύστημα, η κοινότητα από τους χρήστες μπορεί να προσφέρει αυτό το είδος των στοιχείων ρητά. Τα υποκειμενικά στοιχεία μπορούν να είναι η άποψη που προσφέρεται σχετικά με ένα στοιχείο από έναν έμπιστο φίλο. Μια άλλη ανεπάρκεια των content-based συστημάτων είναι η έλλειψη καινοτομίας. Οι collaborative filtering τεχνικές υπερέχουν στον προσδιορισμό της καινοτομίας χρησιμοποιώντας τις συστάσεις άλλων χρηστών και μπορούν να ληφθούν στοιχεία ανόμοια με εκείνα που έχουν ληφθεί στο παρελθόν. Τέλος, τα content-based συστήματα έχουν έλλειψη εκτιμήσεων χρηστών, προκειμένου να παρουσιάσουν τα ενδιαφέροντα του χρήστη. Τα collaborative συστήματα μπορούν να ολοκληρώσουν τις πληροφορίες για έναν χρήστη, βασιζόμενα στην εμπειρία ενός άλλου χρήστη. Τέλος, ο αριθμός των συμμετεχόντων δεν είναι σημαντικός στα content-based συστήματα επειδή δεν εξαρτώνται από τον πληθυσμό. Κατά συνέπεια, τόσο τα content-based, όσο και τα collaborative filtering συστήματα συμβάλλουν το ένα στην αποτελεσματικότητα του άλλου, αποφεύγοντας τους περιορισμούς που αναφέρονται για κάθε σύστημα

και επιτρέποντας σε ένα ολοκληρωμένο σύστημα να επιτύχει και την αξιοπιστία και την εύνοια στις τυχαίες ανακαλύψεις. Διάφοροι ερευνητές έχουν επιβεβαιώσει την υψηλή επίδοση των υβριδικών συστημάτων σε σχέση με τα αμιγή συστήματα.

ΑΝΑΦΟΡΕΣ

- Abdelbar AM, Ragab S, Mitri S (2003) Applying Co-Evolutionary Particle Swarm Optimization to the Egyptian Board Game Seega. In Proceedings of The First Asian-Pacific Workshop on Genetic Programming, (S.B. Cho, N. X. Hoai and Y. Shan editors), 9-15, Canberra, Australia
- Abonyi J., Feil B. and Abraham A. (2005), Computational Intelligence in Data Mining', Informatica: An International Journal of Computing and Informatics, Vol. 29, No. 1, pp. 3-12
- Abraham A., Das S. and Roy S., Swarm Intelligence Algorithms for Data Clustering, pp. 280-313
- Abraham A., Guo H. and Liu H., Swarm Intelligence: Foundations, Perspectives and Applications in Abraham A. et al.: Swarm Intelligence: Foundations, Perspectives and Applications, Studies in Computational Intelligence (SCI) 26, 3–25 (2006)
- Abraham A, Ramos V (2003) Web Usage Mining Using Artificial Ant Colony Clustering and Genetic Programming, 2003 IEEE Congress on Evolutionary Computation (CEC2003), Australia, IEEE Press, ISBN 0780378040, 1384-1391
- Ackerman, M.S. Augmenting the Organizational Memory: A Field Study of Answer Garden, in Proceedings of CSCW'94 (Chapel Hill NC, October 1994), ACM Press, 243-252.
- Ackerman, M.S. and McDonald, D.W. Answer Garden 2: Merging Organizational Memory with Collaborative Help, in Proceedings of CSCW'96 (Boston MA, November 1996), ACM Press, 97-105.
- Admane L, Benatchba K, KoudilM, Siad L, Maziz S (2006) AntPart: an algorithm for the unsupervised classification problem using ants, Applied Mathematics and Computation (<http://dx.doi.org/10.1016/j.amc.2005.11.130>)
- Aggarwal, C.A., Wolf, J.L., Wu, K-L., and Yu, P.S. Horting Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering, in Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1999.
- Amento, B., Hill, W., Terveen, L., Hix, D., and Ju, P. An Empirical Evaluation of User Interfaces for Topic Management of Web Sites, in Proceedings of CHI'99 (Pittsburgh, PA, May 1990), ACM Press, 552-559.
- Avery, C., Resnick, P., and Zeckhauser, R. The Market for Evaluations. American Economic Review 89(3), (1999), 564-584.
- Avgeriou, P., Retails, S., & Skordalakis, M. (2003). An architecture for open learning management systems. In Y. Manolopoulos, S. Evripidou, A.C. Kakas (Eds.) Advances in Informatics. 8th Panhellenic Conference on Informatics. Lectures Notes in Computer Science. LNCS 2563. Berlin: Springer Verlag, pp.183-200.
- Balabanovic, M. and Shoham, Y. Fab: Content-Based, Collaborative Recommendation, in Resnick and Varian (eds.), 66-72
- Barrett, C. Anatomy of a Weblog, Camworld, January 26, 1999. <http://www.camworld.com/journal/rants/99/01/26.html>.
- Barrat A, Weight M (2000) On the properties of small-world network models. The European Physical Journal, 13:547-560
- Blum C (2005) Ant colony optimization: Introduction and recent trends. Physics of Life Reviews, 2, 353–373
- Bentley, P. J. and Corne, D. W. 2001. Creative Evolutionary Systems. Morgan Kaufman Pub.
- Berlangá, A., García, F., Gil, A. "Adaptive Hypermedia Systems: Personalized Education in E-Learning". Educational Virtual Spaces in Practice, The Odiseame Approach, F.J. García, J. García, M. López, R. López and E. Verdú (Eds.), pp. 133-146, Editorial Ariel. ISBN: 84-344-2659-5, 2005.
- Biebricher, C.K., Nicolis, G., Schuster, P., 1995: Self-Organization in the Physico- Chemical and Life Sciences, EU Report 16546.

- Bonabeau E., M. Dorigo & T. Theraulaz, 1999. *From Natural to Artificial Swarm Intelligence*. New York: Oxford University Press.
- Billsus, D. & Pazzani, M. Learning Collaborative Information Filters, in Proceedings of the International Conference on Machine Learning (Madison WI, July 1998), Morgan Kaufmann Publishers.
- Breese, J.S., Heckerman, D., Kadie, C. Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, pp. 43-52, 1998
- Bush, V. As We May Think. *The Atlantic Monthly*, July 1945.
- Card, S.K., Robertson, G.C., and York, W. The WebBook and the Web Forager: An Information Workspace for the World-Wide Web, in Proceedings of CHI'96 (Vancouver BC, April 1996), ACM Press, 111-117.
- Chalmers, M., Rodden, K., and Brodbeck, D. The Order of Things: Activity-Centred Information Access, in Proceedings of 7th International Conference on the World Wide Web, 1998. (Brisbane Australia, April 1998), 359-367.
- Blackwell, T.M. and Bentley, P.J. 2002. Dynamic Search with Charged Swarms. In Proceedings of the Genetic and Evolutionary Computation Conference 2002. New York.
- Cayzer, S. and Aickelin U. 2001. A Recommender System based on the Immune Network.
- Chen Y, Abraham A, (2006) Hybrid Learning Methods for Stock Index Modeling, Artificial Neural Networks in Finance, Health and Manufacturing: Potential and Challenges, J. Kamruzzaman, R. K. Beggs and R. A. Sarker (Eds.), Idea Group Inc. Publishers, USA
- Chen Y, Abraham A (2005) Hybrid Neurocomputing for Detection of Breast Cancer, The Fourth IEEE International Workshop on Soft Computing as Transdisciplinary Science and Technology (WSTST'05), Japan, Springer Verlag, Germany, pp. 884-892
- Chen Y, Peng L, Abraham A (2006) Programming Hierarchical Takagi Sugeno Fuzzy Systems, The 2nd International Symposium on Evolving Fuzzy Systems (EFS2006), IEEE Press
- Chialvo, D.R. and Millonas M.M. How Swarms Build Cognitive Maps. *The Biology of Intelligent Autonomous Agents*, 144, NATO ASI Series, pp. 439-450. 1995
- Chrétien, L. Organisation Spatiale du Matériel Provenant de l'excavation du nid chez *Messor Barbarus* et des Cadavres d'ouvrières chez *Lasius niger* (Hymenopterae: Formicidae) .PhD thesis, Université Libre de Bruxelles, 1996.
- Couzin ID, Krause J, James R, Ruxton GD, Franks NR (2002) Collective Memory and Spatial Sorting in Animal Groups, *Journal of Theoretical Biology*, 218, 1-11
- Cui X, Potok TE (2005) Document Clustering Analysis Based on Hybrid PSO+K-means Algorithm, *Journal of Computer Sciences (Special Issue)*, ISSN 1549-3636, pp. 27-33
- Deerwester, D., Dumais, S.T., Landauer, T.K., Furnas, G.W., and Harshman, R.A. Indexing by Latent Semantic Analysis. *Journal of the Society for Information Science*, 41, 6 (1990), 391-407.
- Delgado, J.A. February 2000. Agent-Based Information Filtering and Recommender Systems on the Internet. PhD thesis, Nagoya Institute of Technology.
- Deneubourg JL, Goss S, Franks N, Franks AS, Detrain C, Chretien L (1991) The dynamics of collective sorting: Robot-like ants and ant-like robots. Proceedings of the First International Conference on Simulation of Adaptive Behaviour: From Animals to Animats, Cambridge, MA: MIT Press, 1, 356-365
- Dall'Asta L, Baronchelli A, Barrat A, Loreto V (2006) Agreement dynamics on smallworld networks. *Europhysics Letters* 15.
- Dorigo M, Blum C (2005) Ant colony optimization theory: A survey. *Theoretical Computer Science*, 344(2-3), 243-278
- Deneubourg J.-L., S. Goss, N. Franks, A. Sendova-Franks, C. Detrain, and L. Chretien. The Dynamic of Collective Sorting Robot-like Ants and Ant-like Robots. *Proc. of 1st Conf. on Simulation of Adaptive Behavior: From Animal to Animats*, edited by J.A. Meyer and S.W. Wilson, pp. 356-365. Cambridge, MA: MIT Press. 1991

- Dorigo M. and G. Di Caro, 1999. The Ant Colony Optimization Meta-Heuristic. In D. Corne, M. Dorigo and F. Glover, editors, *New Ideas in Optimization*, McGraw-Hill, 11-32.
- Dorigo, M., Stützle, T. (2004) *Ant Colony Optimization*. MIT Press.
- Egger F. N. (2001). *Affective Design of E-Commerce User Interfaces: How to Maximise Perceived Trustworthiness*. In *Proceedings of The International Conference on Affective Human Factors Design*. Asean Academic Press, London.
- Dorigo M, Di Caro G, Gambardella LM (1999) Ant algorithms for discrete optimization *Artificial Life*, 5(2), 137-72
- Dorigo M, Gambardella LM (1997) Ant colony system: A cooperative learning approach to the traveling salesman problem. *IEEE Transaction on Evolutionary Computation*, 1(1), 53-66
- Dorigo M, Bonaneau E, Theraulaz G (2000) Ant algorithms and stigmergy, *Future Generation Computer Systems*, 16, 851-871
- Dourish, P. and Bly, S. Portholes: Supporting Awareness in a Distributed WorkGroup, in *Proceedings of CHI'92 (Monterey CA, May 1992)*, ACM Press, 541-547.
- Dourish, P. Where the Footprints Lead: Tracking Down Other Roles for Social Navigation, in Munro, Höök, and Benyon (Eds.), 15-34.
- Eberhart RC, Kennedy J (1995) A new optimizer using particle swarm theory. In *Proceedings of the Sixth International Symposium on Micromachine and Human Science*, Nagoya, Japan, 39-43
- Eberhart RC, Shi Y (2001) Particle swarm optimization: developments, applications and resources. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC)*, Seoul, Korea
- Eberhart RC, Simpson PK, Dobbins RW (1996) *Computational Intelligence PC Tools*, Boston, MA: Academic Press Professional
- Fayyad U, Piatestku-Shapiro G, Smyth P, Uthurusamy R (1996) *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press
- Eberhart, R.C. and Shi, Y. 2001. Particle Swarm Optimization: Developments, Applications and Resources. In *Proceedings of the 2001 Congress on Evolutionary Computation*. vol.1, pp.81-86.
- Flake G (1999) *The Computational Beauty of Nature*. Cambridge, MA: MIT Press
- Fun Y, Chen CY (2005) Alternative KPSO-Clustering Algorithm, *Tamkang Journal of Science and Engineering*, 8(2), 165-174
- Gil A.-B., Garcia-Penalvo F.J.: Learner Course Recommendation in e-Learning Based on Swarm Intelligence, *Journal of Universal Computer Science*, vol. 14, no. 16 (2008), 2737-2755
- Goldberg, D., Nichols, D., Oki, B.M. and Terry, D. Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM*, 35, 12 (December 1992), 51-60.
- Gil, A. and F. Garcia (2003). "E-commerce Recommenders: Powerful Tools for E-business." *Crossroads - The ACM Student Magazine* 10(2): 24-28.
- Gil A.B., García, F.J. (2006) "Recommender Systems in E-Commerce". In *Encyclopedia of Human Computer Interaction*. Idea Group Publishers. Claude Ghaoui Editor. ISBN:1-59140-562-9, 2006.
- Gil, A.B. "E-Commerce y Web-Services: Un Paso Adelante en la Interoperabilidad". In *Tendencias en el Desarrollo de Aplicaciones Web*. F. García Peñalvo y M^a Moreno García (Eds.), Departamento de Informática y Automática de la Universidad de Salamanca, pp. 73-84, 2004.
- Good, N., Schafer, J.B., Konstan, J., Borchers, A., Sarwar, B., Herlocker, J., and Riedl, J., Combining Collaborative Filtering with Personal Agents for Better Recommendations, in *Proceedings of AAAI'99 (July 1999)*.
- Grosan C., Abraham A. and Chis M., (2006) *Swarm Intelligence in Data Mining*, in Abraham A., Grosan C. and Ramos V. (Eds.), *Swarm Intelligence in Data Mining - Studies in Computational Intelligence*, Volume 34, Springer, pp. 17-36.

- Handl J, Knowles J, Dorigo M (2006) Ant-based clustering and topographic mapping *Artificial Life* 12(1)
- Hu X, Shi Y, Eberhart RC (2004) Recent Advances in Particle Swarm, In Proceedings of Congress on evolutionary Computation (CEC), Portland, Oregon, 90-97
- Herlocker, J. Explanations in Recommender Systems, position paper in CHI'99 Workshop Interacting with Recommender Systems, available from <http://www.darmstadt.gmd.de/rec99/schedule.html>.
- Herlocker, J.L., Konstan, J. A. and Riedl, J. 2000. Explaining Collaborative Filtering Recommendations. Proceedings of the ACM 2000 Conference on Computer Supported Cooperative Work.
- Heylighen F. & Gershenson C. (2003): "The Meaning of Self-organization in Computing", IEEE Intelligent Systems, section Trends & Controversies - Self-organization and Information Systems, May/June 2003
- Hill, W.C., Hollan, J.D., Wroblewski, D., and McCandless, T., Edit Wear and Read Wear, in Proceedings of CHI' 92. (Monterey CA, May 1992), ACM Press, 3-9.
- Hill, W.C., Hollan, J.D. History-Enriched Digital Objects: Prototypes and Policy Issues. *The Information Society*, 10, 2 (1994), 139-145.
- Hill, W.C., Stead, L., Rosenstein, M. and Furnas, G. Recommending and Evaluating Choices in a Virtual Community of Use, in Proceedings of CHI'95 (Denver CO, May 1995), ACM Press, 194-201.
- Hill, W. C. and Terveen, L. G. Using Frequency-of-Mention in Public Conversations for Social Filtering. in Proceedings of CSCW'96 (Boston MA, November 1996), ACM Press, 106-112.
- Jasch F, Blumen A (2001) Trapping of random walks on small-world networks. *Physical Review E* 64, 066104
- Jones G, Robertson A, Santimetrovirul C, Willett P (1995) Non-hierarchical document clustering using a genetic algorithm. *Information Research*, 1(1)
- IMS LOM. (2006). Learning Resource Metadata Specification. <http://www.imsglobal.org/metadata/index.html#version1.3>
- Katz, J. Here Come The Weblogs, *Slashdot* May 24, 1999. <http://slashdot.org/features/99/05/13/1832251.shtml>.
- Kautz, H., Selman, B., and Shah, M. The Hidden Web, *AI Magazine*, 18, 2 (Summer 1997), 27 - 36.
- Kennedy J., SWARM INTELLIGENCE, Chapter 6, pp. 187-219
- Kennedy J, Eberhart RC (1995) Particle Swarm Optimization. In Proceedings of IEEE International Conference on Neural Networks, Perth, Australia, IEEE Service Center, Piscataway, NJ, Vol.IV, 1942-1948
- Kennedy J (1997) Minds and cultures: Particle swarm implications. *Socially Intelligent Agents. Papers from the 1997 AAI Fall Symposium*. Technical Report FS-97-02, Menlo Park, CA: AAI Press, 67-72
- Kennedy J (1998) The Behavior of Particles, In Proceedings of 7th Annual Conference on Evolutionary Programming, San Diego, USA
- Kennedy J (1997) The Particle Swarm: Social Adaptation of Knowledge. In Proceedings of IEEE International Conference on Evolutionary Computation, Indianapolis, Indiana, IEEE Service Center, Piscataway, NJ, 303-308
- Kennedy J (1992) Thinking is social: Experiments with the adaptive culture model. *Journal of Conflict Resolution*, 42, 56-76
- Kennedy J, Eberhart R (2001) *Swarm Intelligence*, Morgan Kaufmann Academic Press
- Kennedy J, Mendes R (2002) Population structure and particle swarm performance. In Proceedings of the IEEE Congress on Evolutionary Computation (CEC), 1671-1676
- Kleinberg, J.M. Authoritative Sources in a Hyperlinked Environment, in Proceedings of 1998 ACM-SIAM Symposium on Discrete Algorithms (San Francisco CA, January 1998), ACM Press.
- Konstant, J.; Miller, B.; Maltz, D.; Herlocker, J.; Gordon, L.; Riedl, J. GroupLens: applying collaborative filtering to Usenet news. *Communications of the ACM*, 40(3): 77-87, 1997

- Krause J, Ruxton GD (2002) *Living in Groups*. Oxford: Oxford University Press
- Krohling RA, Hoffmann F, Coelho LS (2004) Co-evolutionary Particle Swarm Optimization for Min-Max Problems using Gaussian Distribution. In *Proceedings of the Congress on Evolutionary Computation 2004 (CEC'2004)*, Portland, USA, volume 1, 959-964
- Kuo RJ, Wang HS, Hu TL, Chou SH (2005) Application of ant K-means on clustering analysis, *Computers & Mathematics with Applications*, Volume 50, Issues 10-12, 1709-1724
- Liu Y, Passino KM (2000) *Swarm Intelligence: Literature Overview*, <http://www.ece.osu.edu/passino/swarms.pdf>
- Lieberman, H. *Autonomous Interface Agents*, in *Proceedings of CHI'97 (Atlanta GA, March 1997)*, ACM Press, 67-74.
- Lorenzi F., Scherer dos Santos D. and Ana L. C. Bazzan, «Negotiation for task allocation among agents in case-base Recommender Systems: a swarm-intelligence approach», 2005, Nineteen International Joint Conference on Artificial Intelligence, Edinburgh, Scotland
- Lovbjerg M, Rasmussen TK, Krink T (2001) Hybrid Particle Swarm Optimiser with Breeding and Subpopulations. *Proc. of the third Genetic and Evolutionary Computation Conference (GECCO-2001)*, volume 1, 469-476
- Lumer ED, Faieta B (1994) Diversity and Adaptation in Populations of Clustering Ants. Clio D, Husbands P, Meyer J and Wilson S (Eds.), *Proceedings of the Third International Conference on Simulation of Adaptive Behaviour: From Animals to Animats 3*, Cambridge, MA: MIT Press, 501-508
- Lumer, E. and B. Faieta. Lumer E.D. and Faieta B. Diversity and Adaptation in Populations of Clustering Ants. *Proc. of SAB'94 – 3rd Conf. on Simulation of Adaptive Behavior: From Animal to Animats*, The MIT Press/Bradford Books, Cambridge, MA. 1994
- Lumer, E. and B. Faieta. *Exploratory Database Analysis via Self-Organization*. Unpublished manuscript, 1995. Referred to in (Bonabeau et al, 1999)
- Major PF, Dill LM (1978) The three-dimensional structure of airborne bird flocks. *Behavioral Ecology and Sociobiology*, 4, 111-122
- Maes, P. *Agents That Reduce Work and Information Overload*. *Communications of the ACM* 37,7, 31-40, July 1994.
- Mackinlay, J.D., Rao, R., and Card, S.K. *An Organic User Interface for Searching Citation Links*, in *Proceedings of CHI'95 (Denver CO, May 1995)*, ACM Press, 67-73.
- Maglio, P.P., Farrell, S., and Barrett, R. *How to Define "Place" on the Web*, in *CHI 2000 Workshop Social Navigation: A Design Approach?*, edited by Höök, K., Munro, A., and Wexelblat, A. Maltz, D. and Ehrlich, K. *Pointing the Way: Active Collaborative Filtering*, in *Proceedings of CHI'95 (Denver CO, May 1995)*, ACM Press, 202-209.
- McDonald, D. and Ackerman, M. *Just Talk to Me: A Field Study of Expertise Location*, in *Proceedings of CSCW'98 (Seattle WA, November 1998)*, ACM Press, 315-324.
- Montaner, M., López, B., de la Rosa, J. Ll., "A Taxonomy of Recommender Agents on the Internet", *Artificial Intelligence Review*, Kluwer Academic Publishers. Volume 19, Issue 4, pp. 285-330. June, 2003
- Mooney, R. J.; Roy, L., *Content-Based Book Recommending Using Learning for Text Categorization*, *Proceedings of the V ACM Conference on Digital Libraries*, San Antonio, USA, pp.195-204, 2000
- Morales, E., García, F., & Barrón, Á. (2006). "Quality Learning Objects Management: A proposal for e-learning Systems" 8th International Conference on Enterprise Information Systems (ICEIS'06). Paphos, Cyprus, 23 - 27 May, 2006.
- Munro, A.J, Höök, K., and Benyon, D (Eds.) *Social Navigation of Information Space*. Springer, 1999

- Merkl D (2002) Text mining with self-organizing maps. Handbook of data mining and knowledge, Oxford University Press, Inc. New York, 903-910
- Moore C, Newman MEJ (2000) Epidemics and percolation in small-world networks. Physics. Review. E 61, 5678-5682
- Newman MEJ, Jensen I, Ziff RM (2002) Percolation and epidemics in a two-dimensional small world, Physics Review, E 65, 021904
- Oliveira LS, Britto AS Jr., Sabourin R (2005) Improving Cascading Classifiers with Particle Swarm Optimization, International Conference on Document Analysis and Recognition (ICDAR 2005), Seoul, South Korea, 570-574
- Omran, M. Particle Swarm optimization methods for pattern Recognition and Image Processing, Ph.D. Thesis, University of Pretoria, 2005
- Omran, M., Salman, A. and Engelbrecht, A. P. Image classification using particle swarm optimization. Proceedings of the 4th Asia-Pacific Conference on Simulated Evolution and Learning 2002 (SEAL 2002), Singapore. pp. 370-374, 2002
- OWL-S 1.2 Pre-Release, March 2006. <http://www.ai.sri.com/daml/services/owl-s/1.2/>
- Paolucci, M., Kawamura, T., Payne, T., and Sycara, K., 2002. Semantic matching of web services capabilities. In First International Semantic Web Conference.
- Paolucci, M., Sycara, K., and Kawamura, T. Delivering Semantic Web Services, 2003. Tech. Report CMU-RI-TR-02-32, Robotics Institute, Carnegie Mellon University.
- Parunak, H. Van Dyke, Brueckner, Sven, 2001. Entropy and Self-Organization in Multi-Agent Systems. Proceedings of the International Conference on Autonomous Agents (Agents 2001), 124-130.
- Paredis J (1994) Steps towards coevolutionary classification neural networks, Artificial Life IV, MIT Press, 359-365
- Partridge BL, Pitcher TJ (1980) The sensory basis of fish schools: relative role of lateral line and vision. Journal of Comparative Physiology, 135, 315-325
- Partridge BL (1982) The structure and function of fish schools. Science American, 245, 90-99
- Pirolli, P., Pitkow, J., and Rao, R. Silk from a Sow's Ear: Extracting Usable Structures from the Web, in Proceedings of CHI'96 (Vancouver BC, April 1996), ACM Press, 118-125.
- Pitkow, J., and Pirolli, P. Life, Death, and Lawfulness on the Electronic Frontier, in Proceedings of CHI'97 (Atlanta GA, March 1997), ACM Press, 383-390.
- Pomeroy P (2003) An Introduction to Particle Swarm Optimization, <http://www.adaptiveview.com/articles/ipsop1.html>
- Raghavan VV, Birchard K (1979) A clustering strategy based on a formalism of the reproductive process in a natural system. Proceedings of the Second International Conference on Information Storage and Retrieval, 10-22
- Ramos V, Muge, F, Pina, P (2002) Self-organized data and image retrieval as a consequence of inter-dynamic synergistic relationships in artificial ant colonies. Soft Computing Systems - Design, Management and Applications, Proceedings of the 2nd International Conference on Hybrid Intelligent Systems, IOS Press, 500-509
- Ramos, V., Ajith Abraham, 2004. Evolving a Stigmergic Self-Organized Data-Mining. In ISDA-04, 4th Int. Conf. on Intelligent Systems, Design and Applications, Budapest, Hungary, ISBN 963-7154-30-2, pp. 725-730, August 26-28.
- Ramos, V., Merelo J. Self-Organized Stigmergic Document Maps: Environment as a Mechanism for Context Learning. Procs. Of AEB'02- 1st Spanish Conference on Evolutionary and Bio-inspired Algorithms, Mérida., pp. 284-293. 2002.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. in Proceedings of CSCW'94 (Chapel Hill NC, October 1994), ACM Press, 175-186.

Resnick, P., and Varian, H.R., guest editors, Communications of the ACM, Special issue on Recommender Systems, 40, 3 (March 1997).

Resnick, P. and Varian, H.R., Recommender Systems, in Resnick and Varian (eds.), 56-58.

Rosenberg, M.J. (2001). E-learning. Strategies for delivering knowledge in the digital age, Mc Graw Hill.

Sarwar, B.; Karypis, G.; Konstan, J.; Riedl, J. Analysis of Recommendation Algorithms for E-Commerce. ACM Conference on Electronic Commerce, pp.158-167, 2000

Schafer, J. B., Konstan, J. A. and Riedl, J. January 2001. E-Commerce Recommendation Applications. Journal of Data Mining and Knowledge Discovery.

Schafer, J. B., Konstan, J. and Riedl, J. 1999. Recommender Systems in E-Commerce. Proceedings of the ACM 1999 Conference on Electronic Commerce.

Selim SZ, Ismail MA (1984) K-means Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality, IEEE Transaction on Pattern Analysis and Machine Intelligence, 6, 81-87

Settles M, Rylander B (2002) Neural network learning using particle swarm optimizers. Advances in Information Science and Soft Computing, 224-226

Shardanand, U., and Maes, P. Social Information Filtering: Algorithms for Automating "Word of Mouth". in Proceedings of CHI'95 (Denver CO, May 1995), ACM Press, 210-217.

Shafer, J.; Konstan, J. A.; Riedl, J. E-Commerce Recommendation Applications. Data Mining and Knowledge Discovery, 5 (1/2), pp. 115-153, 2001

Shardanand, U.; Maes, P. Social Information Filtering: Algorithm for Automating "Word of Mouth". Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems, pp. 210-217, 1995

Shehory, O, Kraus, S. Yadgar, O., 1999. Emergent cooperative goalsatisfaction in large scale automated-agent system. Artificial Intelligence Journal Vol 110(1).

Spiekermann, S.; Paraschiv, C. (2002). Motivating Human-Agent Interaction: Transferring Insights from Behavioral Marketing to Interface Design. Electronic Commerce Research, Vol. 2, pp. 255-285.

Shelokar PS, Jayaraman VK, Kulkarni BD (2004) An ant colony classifier system: application to some process engineering problems, Computers & Chemical Engineering, 28(9), 1577-1584

Shi Y, Krohling RA (2002) Co-evolutionary particle swarm optimization to solving minmax problems. In Proceedings of the IEEE Conference on Evolutionary Computation, Hawaii, 1682-1687

Shi Y, Eberhart RC (1998) A modified particle swarm optimizer. In Proceedings of the IEEE Congress on Evolutionary Computation (CEC), Piscataway, NJ. 69-73

Skopos C, Parsopoulos KE, Patsis PA, Vrahatis MN (2005) Particle swarm optimization: an efficient method for tracing periodic orbits in three-dimensional galactic potential, Mon. Not. R. Astron. Soc. 359, 251-260

Sousa T, Neves A, Silva A (2003) Swarm Optimisation as a New Tool for Data Mining, International Parallel and Distributed Processing Symposium (IPDPS'03), 144b

Sousa T, Silva A, Neves A (2004) Particle Swarm based Data Mining Algorithms for classification tasks, Parallel Computing, Volume 30, Issues 5-6, 767-783

Steinbach M, Karypis G, Kumar V, (2000) A Comparison of Document Clustering Techniques. TextMining Workshop, KDD

Terveen, L.G., Selfridge, P.G., and Long, M.D. Living Design Memory: Framework, Implementation, Lessons Learned. Human-Computer Interaction, 10, 1 (1995), 1-38.

Terveen, L.G., Hill, W.C., and Amento, B. Constructing, Organizing, and Visualizing Collections of Topically Related Web Resources. *ACM Transactions on Computer-Human Interaction* 6,1, 67-94, March 1999.

Terveen L. and Hill W., «Beyond Recommender Systems: Helping People Help Each Other», AT&T Labs – Research, In *HCI In The New Millennium*, Jack Carroll, ed., Addison-Wesley, 2001

Toksari MD (2006) Ant colony optimization for finding the global minimum. *Applied Mathematics and Computation*

Tsai CF, Tsai CW, Wu HC, Yang T (2004) ACODF: a novel data clustering approach for data mining in large databases, *Journal of Systems and Software*, Volume 73, Issue 1, 133-145

Ujjin S, Bentley PJ (2002) Learning User Preferences Using Evolution. In *Proceedings of the 4th Asia-Pacific Conference on Simulated Evolution and Learning*, Singapore

Ujjin S, Bentley PJ (2003) Particle swarm optimization recommender system. *Proceedings of the IEEE Swarm Intelligence Symposium 2003 (SIS 2003)*, Indianapolis, Indiana, USA, 124-131

Valdes J (2004) Building Virtual Reality Spaces for Visual Data Mining with Hybrid Evolutionary-Classical Optimization: Application to Microarray Gene Expression Data. *Proceedings of the IASTED International Joint Conference on Artificial Intelligence and Soft Computing (ASC'2004)*, 713-720

van den Bergh, F. and Engelbrecht, A.P. 2001. Training Product Unit Networks using Cooperative Particle Swarm Optimisers. In *Proceedings of IJCNN 2001*, Washington DC.

Viegas, F.B. and Donath, J.S. Chat Circles, in *Proceedings of CHI'99 (Pittsburgh, PA, May 1990)*, ACM Press, 9-16.

Wexelblat, A. and Maes, P. Footprints: History-Rich Tools for Information Foraging, in *Proceedings of CHI'99 (Pittsburgh PA, May 1990)*, ACM Press, 270-277.

Weng SS, Liu YH (2006) Mining time series data for segmentation by using Ant Colony Optimization, *European Journal of Operational Research*, (<http://dx.doi.org/10.1016/j.ejor.2005.09.001>)

Watts DJ (1999) *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton University Press

Watts DJ, Strogatz SH (1998) Collective dynamics of small-world networks. *Nature*, 393, 440-442

Wu KL, Yang MS (2002) Alternative C-means Clustering Algorithms. *Pattern Recognition*, 35, 2267-2278

Zhao Y, Karypis G (2004) Empirical and Theoretical Comparisons of Selected Criterion Functions for Document Clustering. *Machine Learning*, 55(3), 311-331

Zhao Y. X., *Swarm-Intelligence-Based Optimization Algorithms and Their Applications*

LINKS

<http://www.swarmintelligence.org/SIBook/SI.php#Home>
<http://www.swarmintelligence.org/index.php>
<http://www.engr.iupui.edu/~shi/Coference/psopap4.html#tab>
<http://www.chemoton.org/online.html>
<http://www.deitel.com/ResourceCenters/Web20/RecommenderSystems/tabid/1229/Default.aspx>
<http://www.deitel.com/ResourceCenters/Web20/RecommenderSystems/IntroductiontoRecommenderSystems/tabid/1311/Default.aspx>
<http://users.uom.gr/~mans/papiria/hercma2003.pdf>
<http://hci.dme.uma.pt/courses/socialweb08F/7/p56-resnick.pdf>
http://en.wikipedia.org/wiki/Swarm_intelligence
http://en.wikipedia.org/wiki/Evolutionary_computation
http://en.wikipedia.org/wiki/Recommender_system
<http://www.computelligence.org/sis/2006/SIS05RussEtut.pdf>
<http://www.computelligence.org/sis/2007/?q=node/1>
<http://wi-consortium.org/>
<http://www.ieee-ssci.org/>
<http://www.computelligence.org/sis/2007/>
http://www.iro.umontreal.ca/~aimeur/IJCAI-2005/IJCAI_workshop.htm