

UNIVERSITATEA BABEȘ-BOLYAI
FACULTATEA DE MATEMATICĂ ȘI INFORMATICĂ
UNIVERSITATEA EÖTVÖS LORÁND
FACULTATEA DE INFORMATICĂ

Metode de recunoaștere a formelor bazate pe agenți

Rezumatul tezei

Doctorand: Radu D. Găceanu

Îndrumător UBB: Prof. Dr. Horia F. Pop

Îndrumător ELTE: Conf. Dr. Habil. László Kozma

2012

Mulțumiri

Aș dori să mulțumesc îndrumătorilor mei, Prof. Dr. Horia F. Pop și Conf. Dr. Habil. László Kozma, pentru sugestii și pentru sprijinul acordat în tot acest timp. Le mulțumesc mult tuturor colaboratorilor mei.

Autorul dorește să mulțumească pentru sprijinul financiar acordat prin programele cofinanțate de către Programul Operațional Sectorial Dezvoltarea Resurselor Umane, Contract POSDRU 6/1.5/S/3 “Studii doctorale: prin știință spre societate”.

Lista publicatiilor

- [CDG07] C. Chira, D. Dumitrescu, and **R. D. Găceanu**. Stigmergic agent systems for solving NP-hard problems. *Studia Informatica*, Special Issue KEPT-2007: Knowledge Engineering: Principles and Techniques (June 2007):177–184, June 2007. **(indexed MathSciNet, Zentralblatt MATH, EBSCO Publishing)**.
- [GP10] **R. D. Găceanu** and H. F. Pop. An adaptive fuzzy agent clustering algorithm for search engines. In *MACS2010: Proceedings of the 8th Joint Conference on Mathematics and Computer Science*, pages 185–196. Komarno, Slovakia, 2010. **(indexed MathematicalReviews)**.
- [GP11a] **R. D. Găceanu** and H. F. Pop. A context-aware ASM-based clustering algorithm. *Studia Universitatis Babeş-Bolyai Series Informatica*, LVI(2):55–61, 2011. **(indexed MathSciNet, Zentralblatt MATH, EBSCO Publishing)**.
- [GO11] **R. D. Găceanu** and G. Orbán. Using rsl to describe the stock exchange domain. In *microCAD International Scientific Conference*. University of Miskolc, Hungary, 31 March – 1 April 2011. **(International conference)**.
- [GP11b] **R. D. Găceanu** and H. F. Pop. A fuzzy clustering algorithm for dynamic environments. In *KEPT2011: Knowledge Engineering Principles and Techniques, Selected Papers*, Eds: M. Frentiu, H.F. Pop, S. Motogna, pages 119–130. Babeş-Bolyai University, Cluj-Napoca, Romania, July 4–6 2011. **(ISI — Conference Proceedings Citation Index)**.
- [GP11c] **R. D. Găceanu** and H. F. Pop. An incremental ASM-based fuzzy clustering algorithm. In *Informatics'2011, Slovakia, i'11: Proceedings of the Eleventh International Conference on Informatics, Informatics 2011*, Eds: V. Novitzká, Štefan Hudák, pages 198–204. Slovak Society for Applied Cybernetics and Informatics, Rožňava, Slovakia, November 16–18 2011. **(indexed MathematicalReviews)**.
- [Găc11] **Radu D. Găceanu**. A bio-inspired fuzzy agent clustering algorithm for search engines. *Procedia Computer Science*, 7(0):305 – 307, 2011. Proceedings of the 2nd European Future Technologies Conference and Exhibition 2011 (FET 11), 10.1016/j.procs.2011.09.060. <http://www.sciencedirect.com/science/article/pii/S187705091100620X>. **(ISI — Conference Proceedings Citation Index)**.

- [CCGa] Gabriela Czibula, Istvan Czibula, and **Radu D. Găceanu**. Intelligent data structures selection using neural networks. *Knowledge and Information Systems*, Springer London, pages 1–22. 10.1007/s10115-011-0468-3. **(ISI — Science Citation Index Expanded) IF = 2.0** (2010)
- [GP12] R. D. Găceanu and H. F. Pop. An incremental approach to the set covering problem. *Studia Universitatis Babeş-Bolyai Series Informatica*, LVIII(2), 2012. — under review. **(indexed MathSciNet, Zentralblatt MATH, EBSCO Publishing)**.
- [CCGb] Gabriela Czibula, Istvan Czibula, and **Radu D. Găceanu**. A Support Vector Machine Model For Intelligent Selection of Data Representations. *Applied Soft Computing* — under review. **(ISI — Science Citation Index Expanded) IF = 2.1** (2010)

Cuprins

| | |
|--------------|----|
| Introducere | 6 |
| Concluzii | 11 |
| Bibliografie | 14 |

Contents of the thesis

| | |
|--|-----------|
| Introduction | 6 |
| 1 Theoretical background | 10 |
| 1.1 Data analysis and data mining | 10 |
| 1.2 Pattern recognition | 11 |
| 1.2.1 Cluster analysis | 11 |
| 1.2.2 Classification | 13 |
| 1.3 Soft computing | 19 |
| 1.4 Multi agent interactions | 22 |
| 1.4.1 Direct agent interactions | 22 |
| 1.4.2 Indirect agent interactions | 25 |
| 2 Contributions to NP optimization problems | 28 |
| 2.1 NP-completeness | 28 |
| 2.2 Stigmergic agents | 29 |
| 2.3 Soft agents | 31 |
| 2.3.1 General agent models | 31 |
| 2.3.2 The soft agent model | 33 |
| 2.3.3 Possible extensions and applications | 38 |
| 2.4 A new approach to the set covering problem | 39 |
| 2.4.1 SCP overview | 39 |
| 2.4.2 Incremental SCP | 41 |
| 2.5 Conclusions and future work | 43 |
| 3 New approaches to unsupervised learning | 45 |
| 3.1 Agent-based unsupervised learning | 46 |
| 3.2 ASM-based clustering | 48 |
| 3.2.1 Fuzzy ASM-based clustering | 49 |
| 3.2.2 Context-aware ASM-based clustering | 50 |
| 3.2.3 Case studies | 53 |
| 3.2.4 Discussion | 60 |
| 3.3 Incremental clustering | 61 |
| 3.3.1 General considerations | 61 |

| | | |
|----------|---|------------|
| 3.3.2 | Incremental ASM approach | 63 |
| 3.3.3 | Experiments | 64 |
| 3.4 | Conclusions and future work | 67 |
| 4 | Supervised learning in software development | 69 |
| 4.1 | The problem of dynamic data structure selection | 70 |
| 4.1.1 | Example | 72 |
| 4.1.2 | Experiment | 73 |
| 4.2 | Automatic selection of data representations using ANN | 75 |
| 4.2.1 | Formal aspects | 76 |
| 4.2.2 | Methodology | 77 |
| 4.3 | Experimental evaluation | 82 |
| 4.3.1 | Case study | 82 |
| 4.3.2 | Data collection and pre-processing | 83 |
| 4.3.3 | Testing | 84 |
| 4.3.4 | Discussion | 84 |
| 4.4 | Comparison to related work | 85 |
| 4.5 | Automatic selection of data representations using SVM | 86 |
| 4.5.1 | Overview | 87 |
| 4.5.2 | Formal aspects | 92 |
| 4.5.3 | Methodology | 94 |
| 4.6 | Computational experiments | 97 |
| 4.6.1 | SVM model | 97 |
| 4.6.2 | First case study | 98 |
| 4.6.3 | Second case study | 99 |
| 4.6.4 | Discussion | 102 |
| 4.7 | Comparison to related work | 103 |
| 4.8 | Conclusions and future work | 104 |
| 5 | Conclusions | 105 |
| | Bibliography | 107 |

Introducere

Această lucrare este rezultatul propriei cercetări în domeniul recunoașterii formelor, mai exact în recunoașterea formelor bazate pe agenți, sub îndrumarea Prof. Dr. Horia F. Pop (începând din 2008) și a Conf. Dr. Habil. László Kozma (începând din 2009).

Subiectul cercetării constă în utilizarea mai multor tipuri de agenți software în recunoașterea formelor. Vom investiga în teză utilizarea agenților în probleme de optimizare NP-hard, precum și în analiza de date hibride.

Creșterea rapidă a surselor de date face necesară extragerea și analiza informațiilor din aceste date. Aceste informații și cunoștințe ar putea fi utilizate în diverse aplicații, variind de la detectarea fraudelor, la controlul producției. Analiza datelor poate fi privită ca un pas înainte în evoluția tehnologiei informației. Acesta reprezintă procesul de inspecție, transformare și modelare a datelor, cu scopul de a descoperi modele, asociații și anomalii sprijinind astfel procesul de luare a deciziilor.

Un pas important în analiza datelor este recunoașterea formelor care se ocupă cu atribuirea de etichete unor date de intrare. Clasificarea și *clusteringul* sunt exemple de recunoaștere a formelor. Clasificarea și *clusteringul* pot fi aplicate în multe domenii, cum ar fi în marketing (pentru a găsi grupuri de clienți cu un comportament similar), biologie (clasificarea plantelor și a animalelor date fiind caracteristicile lor), detectarea fraudelor, și clasificarea documentelor.

Clasificarea este procesul prin care se atribuie o etichetă unei date de intrare ceea ce se poate face de exemplu pe baza unui model predefinit. Întrucât eticheta clasei pentru fiecare element din setul de date de *training* este cunoscută, clasificarea este o problemă de învățare supervizată. Pe de altă parte, *clusteringul* este o problemă de învățare nesupervizată și se ocupă cu găsirea unei structuri într-o colecție de date neetichetate. Clasificarea și *clusteringul* împreună cu o descriere generală a analizei datelor sunt prezentate în Capitolul 1.

Atât în clasificare cât și în *clustering* un obiect aparținând aceleiași clase sau *cluster* trebuie să fie similar cu restul elementelor din grup și diferit în raport cu elemente din alte grupuri. Acest lucru implică un grad ridicat de imprecizie, de incertitudine și o modalitate de rezolvare a acestui neajuns este prin utilizarea tehnicilor de soft *computing*. Tehnicile de *soft computing* lucrează cu imprecizia și incertitudinea, în încercarea de a obține robustețe și soluții de cost redus. Acest domeniu multidisciplinar a fost introdus de Lotfi A. Zadeh și scopul său principal este de a dezvolta sisteme inteligente și a rezolva probleme greu de modelat matematic.

Tehnicile de soft computing deschid posibilitatea de a rezolva probleme complexe, pentru care un model matematic nu este disponibil. Mai mult decât atât, introduce concepte umane

precum învățarea în domeniul informaticii. Acest lucru deschide calea pentru construirea de sisteme inteligente și autonome capabile să învețe.

În vederea proiectării sistemelor inteligente și autonome se folosesc agenți software. Un agent este o entitate care poate fi privită ca percepându-și mediul, prin senzori și acționând asupra mediului prin efectori. Un agent care încearcă mereu să-și optimizeze performanța este numit un agent rațional. Agenții au mai multe proprietăți ([SP04, Ser06]), dintre care cea mai interesantă este cea de auto-organizare. Acesta este capacitatea unei entități de a se organiza și de a-și îmbunătăți comportamentul său fără a fi ghidat. Într-un anumit mediu există în general mai mulți agenți care interacționează și formează sisteme multi-agent. Capitolul 1 prezintă mai multe tipuri de interacțiuni într-un sistem multi-agent.

Teza este structurată după cum urmează.

Capitolul 1, **Considerații teoretice**, prezintă domeniul analizei datelor și al analizei datelor bazate pe agenți. Analiza datelor devine din ce în ce mai importantă datorită faptului că ne confruntăm cu tot mai multe date și nevoia de a putea extrage informații relevante din aceste date devine din ce în ce mai necesară. Informațiile relevante obținute prin analiza datelor sunt folosite în diverse aplicații precum detectarea intruziunilor în sisteme, controlul producției sau recunoașterea formelor. Analiza datelor poate fi văzută ca un pas înainte în evoluția tehnologiei informației. Acest capitol prezintă unele dintre cele mai importante probleme în analiza datelor precum *clustering*, clasificare, cât și utilizarea agenților software în acest proces.

Capitolele 2, 3 și 4 conțin abordările noastre originale în domeniul recunoașterii formelor bazată pe agenți. Pentru fiecare abordare propusă subliniem posibilitățile de îmbunătățire și direcțiile viitoare de cercetare. Capitolul subliniază concluziile tezei.

Capitolul 2, **Contribuții în optimizarea problemelor NP**, începe cu o scurtă privire de ansamblu asupra NP completitudinii în Secțiunea 2.1. Restul capitolului este în întregime original și prezintă contribuția noastră în domeniul optimizării problemelor NP, concentrându-se pe două foarte cunoscute probleme NP-dificile: problema comis voiajorului (Travelling Salesman Problem — TSP) și problema acoperirii multimii (Set Covering Problem — SCP). În Secțiunea 2.1 se face o scurtă prezentare a NP completitudinii. În Secțiunea 2.2 problema comis voiajorului este abordată folosind modelul agenților stigmergici. Modelul agenților stigmergici (Stigmergic Agent System — SAS) combină avantajele sistemelor multi-agent (Multi-agent Systems — MAS) și a coloniilor de furnici (Ant Colony Systems — ACS). Interacțiunea cu mediul permite un mecanism general care leagă comportamentul individual de cel al grupului în felul următor: indivizii modifică mediul care la rândul lui modifică felul în care se comportă fiecare individ. Mecanismul SAS folosește mai mulți agenți care interacționează în vederea rezolvării unor probleme prin utilizarea comunicării directe între indivizi și a comunicării indirecte prin intermediul mediului. Algoritmii au fost evaluați pe mai multe seturi de date standard subliniind potențialul metodei. În Secțiunea 2.3 este introdus modelul agenților soft. Un agent soft este un agent inteligent care este capabil să opereze în medii guvernate de imprecizie, incertitudine și aproximare de-a lungul execuției sale ca agent pur reactiv, ca agent orientat spre scop sau

ca ambele. Acest nou model de agent este folosit in Secțiunea 2.4 unde se prezintă o nouă abordare incrementală a problemei acoperirii mulțimii. Experimente pe seturi de date standard sugerează că abordarea este promițătoare. Secțiunea 2.5 subliniază concluziile capitolului si indică direcții de cercetare viitoare.

Capitolul 3, **Noi abordări în învățarea nesupervizată**, începe printr-o scurtă privire de ansamblu asupra diferitelor abordări a custeringului bazat pe agenți în Secțiunea 3.1. Restul capitolului este în întregime original si prezintă contribuția noastră în domeniul clusteringului bazat pe agenți în special în doua direcții principale: clustering static bazat pe agenți ASM și clustering incremental [EKS⁺98, Kam10, LKC02, LLLH10, DL11]. Ne concentrăm asupra dezvoltării de algoritmi de clustering care sunt capabili să descopere date hibride. In Secțiunea 3.1 sunt prezentate câteva metode de clustering bazate pe agenți din literatura de specialitate. Abordarea noastră asupra analizei clusterilor bazată pe agenți este prezentată in Secțiunea 3.2. Datele sunt reprezentate prin agenți plasați într-un grid bidimensional. Agenții se vor grupa in clusteri făcând mutari simple în grid conform unor informații locale, parametrii fiind ajustați adaptiv. Comportamentul este bazate pe ASM (Ant Sleeping Model [CXC04]) unde un agent poate fi fie in stare activa fie in stare inactiva. Pentru a evita blocarea agenților într-un minim local ei sunt capabili să comunice direct între ei. Mai mult decât atât, mutările agenților sunt exprimate prin reguli IF-THEN fuzzy si astfel hibridizarea cu un algoritm clasic de clusterizare nu mai este necesară. Algoritmul de clustering fuzzy bazat pe ASM este prezentat în Secțiunea 3.2.1. In acest model datele care trebuie clusterizate sunt reprezentate prin agenți care sunt capabili sa reacționeze conform schimbărilor din mediu adică numărul agenților din vecinătate. Dar schimbările care pot surveni în date nu sunt tratate în timpul execuției. O extensie la un sistem conștient de context (*context-aware*) ar fi util în multe situații practice. In general, sistemele context-aware ar putea schimba felul in care interacționăm cu mediul — ar putea anticipa nevoile noastre si ne-ar putea asista in luarea deciziilor. Într-un mediu dinamic cunoașterea contextului este indubitabil benefică. Astfel de sisteme ar putea face recomandări relevante si ne-ar putea ajuta în luarea deciziilor. O extensie la o abordare context-aware este prezentată în Secțiunea 3.2.2. Studii de caz pentru ambele abordări incluzând teste pe seturi de date standard [Iri88, Win91] sunt prezentate în Secțiunea 3.2.3. In clusteringul incremental ideea este ca este posibil să consideram pe rand câte o instanța și să o asignăm unuia dintre clusterii deja existenți fără a afecta semnificativ structura deja existentă. Secțiunea 3.3 prezintă abordarea de clustering incrementală bazată pe ASM. În clusteringul incremental numai reprezentanții clusterilor trebuie menținuți în memorie, nu tot setul de date, și deci cerințele pentru spațiul de memorie sunt mici. De fiecare data când o noua instanță este considerată, un algoritm de clustering incremental va încerca practic să o asigneze unuia dintre clusterii deja existenți. Un astfel de proces nu este foarte complex si deci cerințele ca timp de execuție sunt de asemenea mici. Fiind fuzzy, abordarea permite descoperirea de date hibride. Evaluarea experimentală pe seturi de date standard [Iri88, Win91] este prezentată în Secțiunea 3.3.3. Secțiunea 3.4 subliniază concluziile capitolului si indică direcțiile de cercetare ulterioară.

Capitolul 4, **Noi abordări de învățare supervizată în dezvoltarea software**, este în

întregime original și tratează problema selectării dinamice folosind tehnici de învățare supervizată a celei mai potrivite reprezentări a unui tip abstract de date în funcție de contextul de execuție curent a sistemului. În acest sens este propusă o abordare bazată pe rețele neuronale și o abordare bazată pe vectori suport. Selectarea și crearea structurii de date potrivite în implementarea unui tip abstract de date (TAD) poate influența în foarte mare măsură performanța unui sistem. Pentru un dezvoltator de software această problemă este departe de a fi trivială întrucât sunt greu de anticipat toate scenariile de utilizare a unei aplicații. Nu este evident cum s-ar putea selecta o implementare potrivită pentru un tip abstract de date în situația în care modul de accesare este diferit sau chiar impredictibil. Datorită acestui fapt, sistemul software ar trebui să aleagă reprezentarea corespunzătoare la execuție bazându-se pe modul în care tipul abstract de date este folosit. Aceasta selecție dinamică poate fi realizată prin utilizarea tehnicilor de învățare asigurându-se dezvoltarea de sisteme adaptive și complexe. În acest capitol abordăm problema selectării dinamice, folosind tehnici de învățare supervizată, a celei mai potrivite reprezentări a unui tip abstract de date utilizând contextul de execuție curent al sistemului. În acest sens, sunt propuse două abordări: una bazată pe rețele neuronale și una bazată pe vectori suport. Problema considerată a apărut din necesități practice și are o importanță majoră pentru dezvoltatorii software. Utilizarea neadecvată a structurilor de date într-o aplicație duce la degradarea performanței și la consum mare de memorie. Acest lucru poate fi evitat prin selectarea corespunzătoare a structurilor de date care implementează TAD-urile în funcție de natura datelor manipulate. Este prezentată problema selectării dinamice a structurii de date. Este explicat faptul că aceasta este o problemă complexă întrucât orice structură de date este în general mai eficientă pentru anumite operații și mai puțin eficientă pentru altele motiv pentru care o analiză statică pentru alegerea reprezentării celei mai bune nu este potrivită întrucât operațiile care se vor efectua nu pot fi prezise prin analiza statică. Este prezentat un exemplu practic și este efectuat un experiment în vederea motivării abordării noastre. Folosim o rețea neuronală ca prima propunere de utilizare a tehnicilor de învățare supervizată pentru selectarea dinamică a reprezentării a unui tip abstract de date într-un sistem software, selecție bazată pe contextul de execuție curent. De fapt, problema selectării celei mai adecvate implementări a unui tip abstract de date bazându-ne pe contextul de execuție curent, este echivalentă cu prezicerea tipului și a numărului de operații efectuate pe TAD într-un context de execuție dat. Am comparat metoda noastră cu abordările similare din literatura de specialitate. Problema selectării reprezentării datelor este abordată prin folosirea vectorilor suport. Experimentele confirmă o performanță bună a modelului propus accentuând potențialul metodei. Avantajele abordării noastre în comparație cu alte abordări similare din literatura de specialitate sunt totodată evidențiate în acest capitol.

Capitolul , **Concluzii**, conține concluziile tezei.

Contribuțiile originale introduse în teza se afla în Capitolele 2, 3 și 4 și sunt următoarele:

- Un algoritm bazat pe sistemul SAS (Stigmergic Agent System) pentru rezolvarea problemei comis voiajorului (Secțiunea 2.2) [CDG07].
- Un nou model de agenți software: modelul de agenți soft (Secțiunea 2.3) [GP12].

- Un algoritm de clustering incremental pentru rezolvarea problemei acoperirii mulțimii (Secțiunea 2.4) [GP12].
- Evaluarea experimentală a celor doi algoritmi pe seturi de date standard (Secțiunea 2.2 și Secțiunea 2.4) [CDG07, GP12].
- Un algoritm de clustering fuzzy bazat pe ASM (Secțiunea 3.2.1) [GP10, Găc11].
- Un algoritm de clustering fuzzy bazat pe contextul curent (Secțiunea 3.2.2) [GP11a, GP11b].
- Un algoritm fuzzy de clustering incremental (Secțiunea 3.3) [GP11c].
- Evaluarea experimentală a algoritmilor propuși pe seturi de date standard (Secțiunea 3.2.3 și Secțiunea 3.3.3) [GP10, Găc11, GP11a, GP11b, GP11c].
- Descoperirea și analiza datelor hibride (Secțiunea 3.2.3 și Secțiunea 3.3.3) [GP11a, GP11b, GP11c].
- Aplicabilitatea metodelor fuzzy bazate pe ASM în clusterizarea rezultatelor oferite de motoarele de căutare (Secțiunea 3.2.3) [GP10, Găc11].
- O abordare folosind metode de învățare supervizată pentru selecția dinamică a reprezentării tipurilor abstracte de date în timpul execuției unui sistem software (Secțiunea 4.2) [CCGa, CCGb].
- O abordare bazată pe rețele neuronale pentru problema considerată (Secțiunea 4.2.2) [CCGa].
- Evaluarea acurateții tehnicii bazate pe rețele neuronale pe un studiu de caz (Secțiunea 4.3) [CCGa].
- O abordare bazată pe vectori suport pentru problema considerată (Secțiunea 4.5.3) [CCGb].
- Evaluarea acurateții tehnicii bazate pe vectori suport pe un studiu de caz (Secțiunea 4.6) [CCGb].
- O comparație cu alte abordări din literatura de specialitate a avantajelor metodelor de învățare supervizată propuse pentru problema considerată (Secțiunea 4.4 și Secțiunea 4.7) [CCGa, CCGb].

Concluzii

După cum s-a văzut metodele de recunoașterea formelor sunt foarte importante în analiza datelor. Datorită faptului că datele sunt adesea caracterizate de un înalt grad de imprecizie și aproximare, este nevoie de sisteme inteligente și autonome care să poată manipula probleme de complexitate atât de ridicată.

În Capitolul 4 am prezentat un model propriu pentru selecția dinamică a celei mai bune implementări a unui tip abstract de date pe baza contextului de execuție curent dintr-o aplicație software. Pentru a prezice în faza de execuție cea mai adecvată reprezentare, am folosit un model de clasificare bazat pe o rețea neuronală și unul bazat pe vectori suport. Am ilustrat totodată acuratețea ambelor abordări pe studii de caz.

Având în vedere rezultatele prezentate, putem trage concluzia ca abordările introduse aici au următoarele avantaje:

- Acestea sunt generale deoarece pot fi folosite în determinarea reprezentării potrivite pentru orice tip abstract de date și cu un număr aleator de structuri de date care pot fi alese pentru implementarea ADT.
- Reduc timpul de calcul prin selectarea de implementări de structuri de date care oferă o complexitate generală minimă pentru operațiile executate pe un anumit tip abstract de date într-un anumit scenariu de execuție. În consecință eficiența sistemului software de-a lungul evoluției sale este sporită.
- Sunt scalabile deoarece chiar dacă sistemul software considerat este mare, tipurile de date abstracte sunt optimizate local, considerând doar contextul de execuție curent. Dimensiunea contextului de execuție nu depinde de dimensiunea sistemului software.

În Capitolul 3 am prezentat contribuția noastră la clusteringul bazat pe agenți, în special în două direcții principale: clustering static bazat pe ASM și clustering incremental. Ne-am concentrat pe dezvoltarea de algoritmi de clustering care fac posibilă descoperirea și analiza datelor hibride. Algoritmii prezentați sunt bazați pe abordarea ASM adaptivă din [CXC04]. Contribuția majoră este ca în loc să mutăm agenții aleator, îi lăsăm să își aleagă locația cea mai potrivită. Agenții pot comunica direct între ei — asemănător cu abordarea din [CDG07]. În [SCCK04], regulile fuzzy IF-THEN sunt folosite pentru a decide dacă agenții ridică sau lasă jos un obiect. În modelul nostru folosim reguli fuzzy pentru a decide asupra direcției și lungimii unei mutări. Mai mult decât atât, în abordarea context-aware agenții sunt capabili să își adapteze

miscările dacă apar schimbări în mediu. Studii de caz ale acestor abordări sunt efectuate după cum s-a arătat. Pentru a testa algoritmul într-un scenariu real, am considerat seturile de date Iris și Wine [Iri88, Win91]. Experimentele scot în evidență capacitatea metodelor noastre de a descoperi date hibride. Am introdus și un algoritm de clustering incremental. Clusteringul incremental este folosit pentru a procesa fluxuri de date continue sau stream-uri de date și în situații în care forma clusterilor se schimbă în timp. Astfel de algoritmi sunt potriviți în sistemele real-time, rețele de senzori wireless sau streamuri de date deoarece în astfel de sisteme este greu să menținem în memorie întreg setul de date. Algoritmul ia în considerare pe rând câte o instanță și încearcă să o asigneze unuia dintre clusterii deja existenți. Doar reprezentanții clusterului trebuie menținuți în memorie ceea ce duce la o execuție rapidă și fără consum prea mare de memorie. Am văzut în testele de la abordarea incrementală că majoritatea aparentelor erori de clasificare sunt de fapt itemi care au un grad ridicat de apartenență la mai mult de un cluster. Totuși în opinia noastră este din nou clar că avem de-a face cu date hibride. De fapt natura hibridă a datelor este sugerată în [Iri88] și în [Win91] iar acesta este principalul motiv pentru alegerea acestor seturi de date în analiza noastră. Folosind metode fuzzy astfel de caracteristici ale datelor sunt ușor de observat. Existența datelor hibride poate sugera o anumită calitate a datelor.

În Capitolul 2, am prezentat contribuția noastră în problemele de optimizare NP concentrându-ne pe două foarte cunoscute probleme: problema comis voiajorului și problema acoperirii mulțimii. Am prezentat unele considerații generale asupra NP completitudinii. Am abordat problema comis voiajorului folosind modelul agenților stigmergici. Sistemul de agenți stigmergici combină avantajele sistemelor multi-agent cu avantajele sistemelor de colonii de furnici. Comportamentul stigmergic stabilește o legătură între comportamentul indivizilor și cel al coloniei: individul modifică mediul care la rândul său influențează comportamentul altor indivizi. În sistemul de agenți stigmergici, agenții comunică atât în mod direct cât și indirect (stigmergic) în vederea găsirii unei soluții. Algoritmul a fost evaluat pe mai multe seturi de date standard subliniind potențialul metodei. Am introdus modelul agenților soft. Un agent soft este un agent inteligent care trebuie să fie capabil să trateze situații caracterizate prin imprecizie sau aproximare în timpul execuției sale. Am folosit acest nou model într-o abordare incrementală a problemei acoperirii mulțimii. Experimentele pe seturi de date standard sugerează că abordarea este promițătoare.

Ca direcții de cercetare ulterioară intenționăm să îmbunătățim abordările propuse, să extindem evaluarea tehnicilor propuse în teză și să investigăm sau să dezvoltăm alte metode și modele în recunoașterea formelor.

Direcțiile viitoare de cercetare vor fi în zona:

- investigarea altor metaeuristici cu scopul de a identifica alte modele hibride
- folosirea metodelor noastre în rezolvarea altor probleme de optimizare NP
- extinderea metodelor noastre în vederea analizei datelor categoriale
- aplicarea metodelor de clustering incremental în detectarea intruziunilor

- îmbunătățirea modelelor de clasificare propuse pentru selecția dinamică a implementării TAD-urilor prin adăugarea posibilității de adaptare folosind feedback-ul primit în urma unei anumite selecții
- folosirea altor metode de învățare precum hărțile autoorganizate sau alte modele în problema selecției automate a reprezentării TAD-urilor în timpul execuției unui sistem software
- studierea aplicabilității altor metode de învățare precum învățarea semi-supervizată sau învățarea prin întărire în ideea evitării pe cât posibil a supervizării procesului de învățare
- evaluarea tehnicilor noastre în alte studii de caz și sisteme software reale.

Bibliografie

- [AAA03] Noga Alon, Baruch Awerbuch, and Yossi Azar. The online set cover problem. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, STOC '03, pages 100–105, New York, NY, USA, 2003. ACM.
- [ABKS99] Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel, and Jörg Sander. Optics: Ordering points to identify the clustering structure. In *SIGMOD Conference*, pages 49–60, 1999.
- [AM10] Laurent Alfandari and Jérôme Monnot. Approximation of the clustered set covering problem. *Electronic Notes in Discrete Mathematics*, 36:479–485, 2010.
- [AY01] Charu C. Aggarwal and Philip S. Yu. Outlier detection for high dimensional data. In *SIGMOD Conference*, pages 37–46, 2001.
- [AZAY10] Moh'd Belal Al-Zoubi, Al-Dahoud Ali, and Abdelfatah A. Yahya. Fuzzy clustering-based approach for outlier detection. In *Proceedings of the 9th WSEAS international conference on Applications of computer engineering*, ACE'10, pages 192–197, Stevens Point, Wisconsin, USA, 2010. World Scientific and Engineering Academy and Society (WSEAS).
- [BC96] J.E Beasley and P.C Chu. A genetic algorithm for the set covering problem. *European Journal of Operational Research*, 94(2):392 – 404, 1996.
- [Bea] J E Beasley. OR-Library. <http://people.brunel.ac.uk/~mastjjb/jeb/orlib/scpinfo.html>.
- [BG08] Irad Ben-Gal. *Bayesian Networks*. John Wiley & Sons, Ltd, 2008.
- [Bla94a] Betty Blair. Interview with zadeh, creator of fuzzy logic. *Azerbaijan International*, 2(4):46–47, Winter 1994.
- [Bla94b] Betty Blair. Short biographical sketch. *Azerbaijan International*, 2(4):4, Winter 1994.
- [BR03] Christian Blum and Andrea Roli. Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Comput. Surv.*, 35(3):268–308, 2003.

- [BW94] Aart J. C. Bik and Harry A. G. Wijshoff. On automatic data structure selection and code generation for sparse computations. In *Proceedings of the 6th International Workshop on Languages and Compilers for Parallel Computing*, pages 57–75, London, UK, 1994. Springer-Verlag.
- [BW96] Aart J. C. Bik and Harry A. G. Wijshoff. Automatic data structure selection and transformation for sparse matrix computations. *IEEE Trans. Parallel Distrib. Syst.*, 7:109–126, February 1996.
- [CCFM97] Moses Charikar, Chandra Chekuri, Tomás Feder, and Rajeev Motwani. Incremental clustering and dynamic information retrieval. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, STOC '97, pages 626–635, New York, NY, USA, 1997. ACM.
- [CCGa] Gabriela Czibula, Istvan Czibula, and Radu Găceanu. Intelligent data structures selection using neural networks. *Knowledge and Information Systems*, pages 1–22. 10.1007/s10115-011-0468-3.
- [CCGb] Gabriela Czibula, Istvan Czibula, and Radu Găceanu. A support vector machine model for intelligent selection of data representations. *Applied Soft Computing*. under review.
- [CDG07] C. Chira, D. Dumitrescu, and R. D. Găceanu. Stigmergic agent systems for solving NP-hard problems. *Studia Informatica*, Special Issue KEPT-2007: Knowledge Engineering: Principles and Techniques (June 2007):177–184, June 2007.
- [CH96] Tyng-Ruey Chuang and Wen L. Hwang. A probabilistic approach to the problem of automatic selection of data representations. *SIGPLAN Not.*, 31:190–200, June 1996.
- [cir10] Patient contributed image repository. <http://www.pcir.org/>, 2010.
- [CL11] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [CLRS09] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms (3. ed.)*. MIT Press, 2009.
- [CM01] Maria Stella Fiorenzo Catalano and Federico Malucelli. Practical parallel computing. chapter Parallel randomized heuristics for the set covering problem, pages 113–132. Nova Science Publishers, Inc., Commack, NY, USA, 2001.
- [CMP06] C. Chira C. M. Pinteá, P. Pop. Reinforcing ant colony system for the generalized traveling salesman problem. In *Volume of Evolutionary Computing, International Conference Bio-Inspired Computing - Theory and Applications (BIC-TA)*, pages 245–252, New York, NY, USA, September 18–22, 2006. Wuhan, China.

- [CMR⁺07] Oscar Castillo, Patricia Melin, Oscar Montiel Ross, Roberto Sepveda Cruz, Witold Pedrycz, and Janusz Kacprzyk. *Theoretical Advances and Applications of Fuzzy Logic and Soft Computing*. Springer Publishing Company, Incorporated, 1st edition, 2007.
- [coNop] A compendium of NP optimization problems. <http://www.nada.kth.se/viggo/problemelist/compendium.html>.
- [CSM⁺11] Broderick Crawford, Ricardo Soto, Eric Monfroy, Fernando Paredes, and Wenceslao Palma. A hybrid ant algorithm for the set covering problem. *International Journal of the Physical Sciences*, 6(19):4667–4673, September 16 2011.
- [CST00] Nello Cristianini and John Shawe-Taylor. *An introduction to support Vector Machines: and other kernel-based learning methods*. Cambridge University Press, New York, NY, USA, 2000.
- [CXC04] L. Chen, X. H. Xu, and Y. X. Chen. An adaptive ant colony clustering algorithm. In *Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on, Vol. 3*, pages 1387–1392, 2004.
- [DB05] Marco Dorigo and Christian Blum. Ant colony optimization theory: A survey. *Theor. Comput. Sci.*, 344(2-3):243–278, 2005.
- [DDC99] Marco Dorigo and Gianni Di Caro. *The ant colony optimization meta-heuristic*, pages 11–32. McGraw-Hill Ltd., UK, Maidenhead, UK, England, 1999.
- [DGF⁺91] J. L. Deneubourg, S. Goss, N. Franks, A. Sendova-Franks, C. Detrain, and L. Chretien. The dynamic of collective sorting robot-like ants and ant-like robots. In *SAB90 - 1st Conf. On Simulation of Adaptive Behavior: From Animals to Animats*, pages 356–365. MIT Press, 1991.
- [DICiM11] Digital Imaging and Communications in Medicine. <http://medical.nema.org/>, 2011.
- [DL11] Steven Simske Dalong Li. Training set compression by incremental clustering. *Journal of Pattern Recognition Research*, 6:56–64, 2011.
- [Dor07] M. Dorigo. Ant colony optimization. *Scholarpedia*, 2(3):1461, 2007.
- [DP95] Dan Dumitrescu and Horia Florin Pop. Degenerate and non-degenerate convex decomposition of finite fuzzy partitions — I. *Fuzzy Sets and Systems*, 73:365–376, 1995.
- [DP98] Dan Dumitrescu and Horia Florin Pop. Degenerate and non-degenerate convex decomposition of finite fuzzy partitions — II. *Fuzzy Sets and Systems*, 96:111–118, 1998.

- [DS04] M. Dorigo and T. Stützle. *Ant Colony Optimization*. MIT Press, 2004.
- [EKS⁺98] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Michael Wimmer, and Xiaowei Xu. Incremental clustering for mining in a data warehousing environment. In *Proceedings of the 24rd International Conference on Very Large Data Bases, VLDB '98*, pages 323–333, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.
- [EpKXSX96] Martin Ester, Hans peter Kriegel, Jrg S, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press, 1996.
- [Est09] Martin Ester. Density-based clustering. In *Encyclopedia of Database Systems*, pages 795–799. 2009.
- [Etz96] O. Etzioni. Moving up the information food chain: deploying softbots on the world wide web. In *Proceedings of the 13rd national Conference on Artificial Intelligence (AAAI- 96)*, pages 4–8. Portland, OR, 1996.
- [fIPA] Foundation for Intelligent Physical Agents. <http://www.fipa.org/>.
- [FLM97] T. Finin, Y. Labrou, and J. Mayfield. *Kqml as an Agent Communication Language*. Software Agents, B.M. Jeffrey, MIT Press, 1997.
- [FMPS00] Kilian Foth, Wolfgang Menzel, Horia Florin Pop, and Ingo Schröder. An experiment on incremental analysis using robust parsing techniques. In *The 18th International Conference on Computational Linguistics*, pages 1026–1030. Universität des Saarlandes, Saarbrücken, Germany, July-August 2000.
- [For10] M Forina. <http://archive.ics.uci.edu/ml>, 2010.
- [Găc11] Radu D. Găceanu. A bio-inspired fuzzy agent clustering algorithm for search engines. *Procedia Computer Science*, 7(0):305 – 307, 2011. Proceedings of the 2nd European Future Technologies Conference and Exhibition 2011 (FET 11).
- [GC10] Serge Guillaume and Brigitte Charnomordic. Interpretable fuzzy inference systems for cooperation of expert knowledge and data in agricultural applications using fispro. In *FUZZ-IEEE*, pages 1–8, 2010.
- [Gei93] S. Geisser. *Predictive inference: an introduction*. Monographs on statistics and applied probability. Chapman & Hall, 1993.
- [GKP02] Mihaela Gordan, Constantine Kotropoulos, and Ioannis Pitas. A support vector machine-based dynamic network for visual speech recognition applications. *EURASIP J. Appl. Signal Process.*, 2002:1248–1259, January 2002.

- [GKS09] Betsy George, James M. Kang, and Shashi Shekhar. Spatio-temporal sensor graphs (stsg): A data model for the discovery of spatio-temporal patterns. *Intell. Data Anal.*, 13(3):457–475, 2009.
- [Glo89] Fred Glover. Tabu search - part I. *INFORMS Journal on Computing*, 1(3):190–206, 1989.
- [Glo90] Fred Glover. Tabu search - part II. *INFORMS Journal on Computing*, 2(1):4–32, 1990.
- [GO11] R. D. Găceanu and G. Orbán. Using rsl to describe the stock exchange domain. In *microCAD International Scientific Conference*. University of Miskolc, Hungary, 31 March – 1 April 2011.
- [Gol89] David E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1989.
- [GP08] Darwin Gouwanda and S. G. Ponnambalam. Evolutionary search techniques to solve set covering problems. In *World Academy of Science, Engineering and Technology*, pages 20–26. WASET, March 2008.
- [GP10] R. D. Găceanu and H. F. Pop. An adaptive fuzzy agent clustering algorithm for search engines. In *MACS2010: Proceedings of the 8th Joint Conference on Mathematics and Computer Science*, pages 185–196. Komarno, Slovakia, 2010.
- [GP11a] R. D. Găceanu and H. F. Pop. A context-aware ASM-based clustering algorithm. *Studia Universitatis Babes-Bolyai Series Informatica*, LVI(2):55–61, 2011.
- [GP11b] R. D. Găceanu and H. F. Pop. A fuzzy clustering algorithm for dynamic environments. In *KEPT2011: Knowledge Engineering Principles and Techniques, Selected Papers, Eds: M. Frentiu, H.F. Pop, S. Motogna*, pages 119–130. Babes-Bolyai University, Cluj-Napoca, Romania, July 4–6 2011.
- [GP11c] R. D. Găceanu and H. F. Pop. An incremental ASM-based fuzzy clustering algorithm. In *Informatics'2011, Slovakia, i'11: Proceedings of the Eleventh International Conference on Informatics, Informatics 2011, Eds: V. Novitzká, Štefan Hudák*, pages 198–204. Slovak Society for Applied Cybernetics and Informatics, Rožňava, Slovakia, November 16–18 2011.
- [GP12] R. D. Găceanu and H. F. Pop. An incremental approach to the set covering problem. *Studia Universitatis Babes-Bolyai Series Informatica*, LVIII(2), 2012.
- [Har75] J.A. Hartigan. *Clustering algorithms*. Wiley series in probability and mathematical statistics. Applied probability and statistics. Wiley, 1975.

- [HCL00] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. A practical guide to support vector classification, 2000.
- [HFH⁺09] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: an update. *SIGKDD Explorations*, 11(1):10–18, 2009.
- [HK06] Jiawei Han and Micheline Kamber. *Data Mining: Concepts and Techniques, 2nd ed.* Morgan Kaufmann, 2006.
- [HKT01] J. Han, M. Kamber, and A. K. H. Tung. *Spatial Clustering Methods in Data Mining: A Survey.* Taylor and Francis, 2001.
- [HL0] Health Level 7. www.hl7.org/, 0.
- [hp10] DICOM home page. <ftp://medical.nema.org/medical/dicom/datasets/>, 2010.
- [HSP08] Samer Hassan, Mauricio Salgado, and Juan Pavón. Friends forever: Social relationships with a fuzzy agent-based model. In *HAIS*, pages 523–532, 2008.
- [HZK⁺09] Pari Delir Haghighi, Arkady B. Zaslavsky, Shonali Krishnaswamy, Mohamed Medhat Gaber, and Seng Wai Loke. Context-aware adaptive data stream mining. *Intell. Data Anal.*, 13(3):423–434, 2009.
- [Ins] MP-TESTDATA The TSPLIB Symmetric Traveling Salesman Problem Instances. <http://elib.zib.de/pub/mp-testdata/tsp/tsplib/tsp/index.html>.
- [Iri88] Machine Learning Repository Iris. <http://archive.ics.uci.edu/ml/datasets/iris>, 1988.
- [Kam10] A. Kamble. Incremental clustering in data mining using genetic algorithm. *International Journal of Computer Theory and Engineering*, 2(3):1793–8201, 2010.
- [KH01] Gregor Kiczales and Erik Hilsdale. Aspect-oriented programming. *SIGSOFT Softw. Eng. Notes*, 26:313–, September 2001.
- [KHK99] George Karypis, Eui-Hong Han, and Vipin Kumar. Chameleon: Hierarchical clustering using dynamic modeling. *IEEE Computer*, 32(8):68–75, 1999.
- [KJV83] Scott Kirkpatrick, D. Gelatt Jr., and Mario P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.
- [KKvdSS96] Ben Krse, Ben Krose, Patrick van der Smagt, and Patrick Smagt. An introduction to neural networks, 1996.
- [Kot07] Sotiris B. Kotsiantis. Supervised machine learning: A review of classification techniques. *Informatika (Slovenia)*, 31(3):249–268, 2007.

- [KS05] Vikas Kumar and Marta Schuhmacher. Fuzzy uncertainty analysis in system modelling. In *European Symposium on Computer Aided Process Engineering 15 L. Puigjaner and A. Espua (Editors)*, 2005.
- [KS08] P. R. Kumar K. Sreelakshmi. Performance evaluation of short term wind speed prediction techniques. *IJCSNS International Journal of Computer Science and Network Security*, 8(8):162–169, 2008.
- [KTL11] Suzan Köknar-Tezel and Longin Jan Latecki. Improving svm classification on imbalanced time series data sets with ghost points. *Knowl. Inf. Syst.*, 28:1–23, July 2011.
- [KY95] GEORGE J. Klir and BO Yuan. *FUZZY SETS AND FUZZY LOGIC Theory and Applications*. Prentice Hall, 1995.
- [LB05] Kristopher R. Linstrom and A. John Boye. A neural network prediction model for a psychiatric application. In *Proceedings of the Sixth International Conference on Computational Intelligence and Multimedia Applications*, pages 36–40, Washington, DC, USA, 2005. IEEE Computer Society.
- [LF94] E. Lumer and B. Faieta. Diversity and adaptation in populations of clustering ants. In *J.-A.Meyer, S.W.Wilson(Eds.), Proceedings of the Third International Conference on Simulation of Adaptive Behavior: From Animats, Vol.3*, pages 501–508. MIT Press/Bradford Books,Cambridge, MA, 1994.
- [LKC02] Kyung-Soon Lee, Kyo Kageura, and Key-Sun Choi. Implicit ambiguity resolution using incremental clustering in korean-to-english cross-language information retrieval. In *Proceedings of the 19th international conference on Computational linguistics - Volume 1, COLING '02*, pages 1–7, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.
- [LLLH10] Zhenhui Li, Jae-Gil Lee, Xiaolei Li, and Jiawei Han. Incremental clustering for trajectories. In Hiroyuki Kitagawa, Yoshiharu Ishikawa, Qing Li, and Chiemi Watanabe, editors, *Database Systems for Advanced Applications*, volume 5982 of *Lecture Notes in Computer Science*, pages 32–46. Springer Berlin / Heidelberg, 2010.
- [Low78] James R. Low. Automatic data structure selection: an example and overview. *Commun. ACM*, 21:376–385, May 1978.
- [LR76] James Low and Paul Rovner. Techniques for the automatic selection of data structures. In *Proceedings of the 3rd ACM SIGACT-SIGPLAN symposium on Principles on programming languages, POPL '76*, pages 58–67, New York, NY, USA, 1976. ACM.

- [MA75] E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7(1):1–13, 1975.
- [Mar09] Stephen Marsland. *Machine Learning: An Algorithmic Perspective*. Chapman & Hall/CRC, 1st edition, 2009.
- [Mat] Fuzzy Logic Toolbox MathWorks. <http://www.mathworks.com/help/toolbox/fuzzy/fp351dup8.html>.
- [Mit97] Thomas M. Mitchell. *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1 edition, 1997.
- [MK91] S. Miyake and F. Kanaya. A neural network approach to a bayesian statistical decision problem. *IEEE Trans. Neural Networks*, 2:538–540, 1991.
- [Mou01] D.M. Mount. Lecture notes CMSC 420, Data Structures, 2001.
- [NGS11] K Venkatramaiah Deepak P C Navneet Goyal, Poonam Goyal and Sanoop P S. An efficient density based incremental clustering algorithm in data warehousing environment. In *2009 International Conference on Computer Engineering and Applications IPCSIT vol.2 (2011), IACSIT Press, Singapore*, pages 482 – 486, 2011.
- [Ngu10] Nam Nguyen. A new svm approach to multi-instance multi-label learning. In *Proceedings of the 2010 IEEE International Conference on Data Mining, ICDM '10*, pages 384–392, Washington, DC, USA, 2010. IEEE Computer Society.
- [NMM06] Giuseppe Narzisi, Venkatesh Mysore, and Bud Mishra. Multi-objective evolutionary optimization of agent-based models: An application to emergency response planning. In *Computational Intelligence*, pages 228–232, 2006.
- [oPDsf10] Washington State University College of Pharmacy DICOM sample files. <http://info.betaustur.org/>, 2010.
- [Osi10] Advanced Imaging in 3D/4D/5D Sample DICOM Image Sets Osirix. <http://pubimage.hcuge.ch:8080/>, 2010.
- [OSM] ObjectWeb: Open Source Middleware. <http://asm.objectweb.org/>.
- [PS82] Christos H. Papadimitriou and Kenneth Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, 1982.
- [PS96] Horia Florin Pop and Costel Sârbu. A new fuzzy regression algorithm. *Anal. Chem.*, 68:771–778, 1996.

- [PSHD96] Horia Florin Pop, Costel Sârbu, Ossi Horowitz, and Dan Dumitrescu. A fuzzy classification of the chemical elements. *J. Chem. Inf. Comput. Sci.*, 36:465–482, 1996.
- [RHW86] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. *Learning internal representations by error propagation*, pages 318–362. MIT Press, Cambridge, MA, USA, 1986.
- [RiplsDis10] J.-P.: Signal Roux and image processing lab sample DICOM image sets. <http://www.creatis.insa-lyon.fr/~jpr/public/gdcm/gdcmsampledata/>, 2010.
- [RKY10] Vikas C. Raykar, Balaji Krishnapuram, and Shipeng Yu. Designing efficient cascaded classifiers: tradeoff between accuracy and cost. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, pages 853–860, New York, NY, USA, 2010. ACM.
- [RN02] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2002.
- [Roj96] Raúl Rojas. *Neural networks: a systematic introduction*. Springer-Verlag New York, Inc., New York, NY, USA, 1996.
- [Rov78] P. Rovner. *Automatic representation selection for associative data*. Managing Requirements Knowledge, International Workshop, 1978.
- [SB98] Richard S. Sutton and Andrew G. Barto. *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, USA, 1st edition, 1998.
- [SC08] Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Springer Publishing Company, Incorporated, 1st edition, 2008.
- [SCCK04] S. Schockaert, M. De Cock, C. Cornelis, and E. E. Kerre. Fuzzy ant based clustering. In *Ant Colony Optimization and Swarm Intelligence, 4th International Workshop (ANTS 2004), LNCS 3172*, pages 342–349, 2004.
- [sciom11] Open source clinical image and object management. <http://www.dcm4che.org/>, 2011.
- [SdLFdCG09] Eduardo J. Spinosa, André Carlos Ponce de Leon Ferreira de Carvalho, and João Gama. Novelty detection with application to data streams. *Intell. Data Anal.*, 13(3):405–422, 2009.
- [Ser06] Gabriela Serban. *Sisteme Mutiagent in inteligenta artificiala distribuita. Arhitecturi si aplicatii*. Ed. Risoprint, Cluj-Napoca, 2006.
- [sf] Apache software foundation. <http://www.apache.org/>.

- [SGT⁺02] Ingo Schröder, K. Giannakoglou, D. Tsahalis, J. Periaux, K. Papailiou, T. Fogarty (eds, Horia F. Pop, A Fachbereich Informatik, Wolfgang Menzel, and Kilian A. Foth. Learning weights for a natural language grammar using genetic algorithms, 2002.
- [SH07] Mark D. Skowronski and John G. Harris. Automatic speech recognition using a predictive echo state network classifier. *Neural Networks*, 20(3):414 – 423, 2007. Echo State Networks and Liquid State Machines.
- [SK99] Panu Somervuo and Teuvo Kohonen. Self-organizing maps and learning vector quantization for feature sequences. *Neural Process. Lett.*, 10:151–159, October 1999.
- [SP00] Costel Sârbu and Horia Florin Pop. Fuzzy clustering analysis of the first 10 meic chemicals. *Chemosphere*, 40:513–520, 2000.
- [SP04] Gabriela Serban and Horia Florin Pop. *Tehnici de Inteligența Artificială. Abordări bazate pe Agenți Inteligenți*. Ed. Mediamira, Cluj-Napoca, 2004.
- [Spe87] C. Spearman. The proof and measurement of association between two things. By C. Spearman, 1904. *The American journal of psychology*, 100(3-4):441–471, 1987.
- [SS01] Bernhard Scholkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001.
- [SSS79] Edmond Schonberg, Jacob T. Schwartz, and Micha Sharir. Automatic data structure selection in setl. In *Proceedings of the 6th ACM SIGACT-SIGPLAN symposium on Principles of programming languages*, POPL '79, pages 197–210, New York, NY, USA, 1979. ACM.
- [SSS81] Edmond Schonberg, Jacob T. Schwartz, and Micha Sharir. An automatic technique for selection of data representations in setl programs. *ACM Trans. Program. Lang. Syst.*, 3:126–143, April 1981.
- [Ste90] Luc Steels. Components of expertise. *AI Magazine*, 11(2):28–49, 1990.
- [TD10] Balint Takacs and Yiannis Demiris. Spectral clustering in multi-agent systems. *Knowl. Inf. Syst.*, 25:607–622, December 2010.
- [Vap00] V.N. Vapnik. *The nature of statistical learning theory*. Statistics for engineering and information science. Springer, 2000.
- [VSP09] K. R. Venugopal, K. G. Srinivasa, and L. M. Patnaik. *Soft Computing for Data Mining Applications*. Springer Publishing Company, Incorporated, 1st edition, 2009.

- [Wat89] C. J. C. H. Watkins. *Learning from Delayed Rewards*. PhD thesis, Cambridge University, Cambridge, England, 1989.
- [WB01] D.A. Watt and D.F. Brown. *Java collections: an introduction to abstract data types, data structures, and algorithms*. John Wiley, 2001.
- [WD92] Christopher J. C. H. Watkins and Peter Dayan. Technical note q-learning. *Machine Learning*, 8:279–292, 1992.
- [Win91] Machine Learning Repository Wine. <http://archive.ics.uci.edu/ml/datasets/wine>, 1991.
- [WLZ00] G Wahba, Y Lin, and H Zhang. Generalized approximate cross validation for support vector machines, or, another way to look at margin-like quantities. *Advances in large margin classifiers*, (1006):297309, 2000.
- [Woo99] Michael Wooldridge. *Intelligent Agents, An Introduction to Multiagent Systems*. Ed. G. Weiss, 1999.
- [Woo09] Michael J. Wooldridge. *An Introduction to MultiAgent Systems (2. ed.)*. Wiley, 2009.
- [WYM97] Wei Wang, Jiong Yang, and Richard R. Muntz. Sting: A statistical information grid approach to spatial data mining. In *VLDB*, pages 186–195, 1997.
- [Yel03] D. M. Yellin. Competitive algorithms for the dynamic selection of component implementations. *IBM Syst. J.*, 42:85–97, January 2003.
- [Zad65] Lotfi Askar Zadeh. Fuzzy sets. *Inf. Control*, 8:338–353, 1965.
- [Zad94] Lotfi A. Zadeh. Fuzzy logic, neural networks, and soft computing. *Commun. ACM*, 37:77–84, March 1994.
- [Zad97] Lotfi A. Zadeh. The roles of fuzzy logic and soft computing in the conception, design and deployment of intelligent systems. In *Software Agents and Soft Computing*, pages 183–190, 1997.
- [Zad02] Lotfi A. Zadeh. Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. *Journal of Statistical Planning and Inference*, 105(1):233 – 264, 2002. Imprecise Probability Models and their Applications.
- [Zad08] Lotfi Askar Zadeh. Is there a need for fuzzy logic. *Information Sciences*, 178(13):2751–2779, July 2008.
- [ZDYZ11] Xingquan Zhu, Wei Ding, Philip Yu, and Chengqi Zhang. One-class learning and concept summarization for data streams. *Knowledge and Information Systems*, 28:523–553, 2011. 10.1007/s10115-010-0331-y.

- [Zha00] G. P. Zhang. Neural networks for classification: a survey. *IEEE Trans. Systems, Man and Cybernetics*, 30(4):451–462, November 2000.
- [ZL10] Zhi-Hua Zhou and Ming Li. Semi-supervised learning by disagreement. *Knowl. Inf. Syst.*, 24:415–439, September 2010.
- [ZRL97] Tian Zhang, Raghu Ramakrishnan, and Miron Livny. Birch: A new data clustering algorithm and its applications. *Data Min. Knowl. Discov.*, 1(2):141–182, 1997.