



UNIVERSITATEA BABEȘ-BOLYAI, CLUJ-NAPOCA, ROMANIA
FACULTATEA DE MATEMATICĂ ȘI INFORMATICĂ

INSTITUT NATIONAL DES SCIENCES APPLIQUÉES, ROUEN, FRANCE,
LABORATOIRE D 'INFORMATIQUE, DE TRAITEMENT DE L'INFORMATION ET
DES SYSTÈMES

Modele dinamice de instruire automată pentru clasificare supervizată și nesupervizată

Rezumat

Student doctorand:

Adela-Maria SÎRBU (căs. RUS)

Conducători științifici:

Profesor Dr. **Gabriela CZIBULA** – *Universitatea Babeș-Bolyai, România*

Profesor Dr. **Abdelaziz BENSRAIR** – *INSA de Rouen, France*

Iunie 2016

Mulțumiri

În primul rând, doresc să-mi exprim profunda recunoștință față de conducătorii mei științifici Prof. Dr. Gabriela Czibula și Prof. Dr. Abdelaziz Bensrhair, pentru buna îndrumare, răbdare, suport științific și financiar pe parcursul acestor ani de doctorat. Fără ajutorul lor nu as fi putut finaliza această teză.

Doresc să mulțumesc în mod special dnei Conf. Dr. Alexandrina Rogozan, care m-a ghidat de asemenea în munca de cercetare, m-a încurajat și a facut posibilă colaborarea cu INSA.

În al doilea rând, doresc să-mi exprim recunoștința față de juriul care a acceptat să-mi examineze teza Prof. Dr. Horia F. Pop (Universitatea Babeș-Bolyai), Prof. Dr. Daniela Zaharie (Universitatea de Vest din Timisoara), Prof. Dr. Fawzi Nashashibi (INRIA, Franta) și Prof. Dr. Laurent Vercouter (INSA de Rouen).

Facând parte din două universități, am avut oportunitatea de a cunoaște persoane valoroase care m-au ajutat atât profesional cât și personal. Doresc să-i mulțumesc dnei Conf. Dr. Laura Dioșan pentru îndrumarea ei permanentă încă din al doilea an de facultate. Aș vrea să le mulțumesc de asemenea colegilor mei de la Universitatea Babeș-Bolyai: Iuliana, Zsuzsanna și Gabriel și de la INSA: Alina, Vannee, Bassem, Fabian pentru ajutorul și prietenia lor. A fost minunat să fac parte dintr-o așa echipa.

Nu în ultimul rând doresc să mulțumesc soțului meu Florin, familiilor mele (Sîrbu and Rus) și prietenilor mei pentru că au crezut în mine, pentru îndelunga lor răbdare și ajutorul lor necondiționat pe parcursul acestor ani dificili.

Cuprins

Introducere	5
Concluzii	7
Bibliografie	9

Lista publicațiilor

Publicații în ISI Web of Knowledge

Publicații în ISI Science Citation Index Expanded

1. Gabriela Czibula, Istvan-Gergely Czibula, **Adela Sîrbu** and Gabriel Mircea A novel approach to adaptive relational association rule mining. *Applied Soft Computing* published by *Elsevier*, volume 36, pp. 519–533, 2015
2. Bassem Besbes, Alexandrina Rogozan, **Adela-Maria Rus (Sîrbu)**, Abdelaziz Bensrhair and Alberto Broggi. Pedestrian Detection în Far-Infrared Daytime Images Using a Hierarchical Codebook of SURF. *Sensors*, volume 15, no. 4, pp. 8570-8594, 2015, doi:10.3390/s150408570

Publicații în ISI Conference Proceedings Citation Index

3. **Adela-Maria Sîrbu**, Gabriela Czibula and Maria-Iuliana Bocicor. Dynamic Clustering of Gene Expression Data Using a Fuzzy Approach. *16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, SYNASC 2014, Timisoara, Romania, September 22-25, pp. 220-227, 2014
4. **Adela-Maria Sîrbu**, Alexandrina Rogozan, Laura Dioșan and Abdelaziz Bensrhair. Pedestrian Recognition by Using a Kernel-Based Multi-modality Approach. *16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, SYNASC 2014, Timisoara, Romania, September 22-25, pp. 258-263, 2014

Articole publicate în jurnale internaționale și volume ale unor conferințe internaționale

5. **Adela Sîrbu** and Maria-Iuliana Bocicor. A dynamic approach for hierarchical clustering of gene expression data. *Proceedings of IEEE International Conference on Intelligent Computer Communication and Processing*, Cluj-Napoca, Romania, pp. 3-6, 2013.

6. **Adela-Maria Rus (Sîrbu)**, Alexandrina Rogozan, Laura Dioşan and Abdelaziz Bensrhair. Pedestrian recognition using a dynamic modality fusion approach. *Proceedings of IEEE International Conference on Intelligent Computer Communication and Processing*, Cluj-Napoca, Romania, September 3-5, pp. 393 - 400, 2015
7. **Adela-Maria Rus (Sîrbu)**, Alexandrina Rogozan, Laura Dioşan and Abdelaziz Bensrhair. Pedestrian recognition by using a dynamic modality selection approach. *Proceedings of IEEE 18th International Conference on Intelligent Transportation Systems*, Las Palmas, Spain, September 15-18, pp. 1862 - 1867, 2015
8. Maria-Iuliana Bocicor, **Adela-Maria Sîrbu** and Gabriela Czibula. Dynamic core based clustering of gene expression data. *International Journal of Innovative Computing, Information and Control*, volume 10, no. 3, pp. 1051-1069, 2014
9. **Adela Sîrbu**. A study on dynamic clustering of gene expression data. *Studia Universitatis "Babes-Bolyai", Informatica*, Romania, Vol. LIX, Number 1, pp. 16-27, 2014.
10. Anca Andreica, Laura Dioşan, Radu Găceanu and **Adela Sîrbu**. Pedestrian recognition by using kernel descriptors. *Studia Universitatis "Babes-Bolyai", Informatica*, Romania, Vol. LVII, Number 2, pp. 77-89, 2013.

Introducere

Această teză de doctorat este rezultatul cercetării mele în domeniul aplicării modelelor dinamice de instruire automată pentru rezolvarea problemelor de clasificare supervizată și nesupervizată. Cercetarea a fost începută în anul 2012, sub îndrumarea coordonatorilor științifici Prof. Dr. Gabriela Czibula și Prof. Dr. Abdelaziz Bensrhair.

Instruirea automată este o ramură a inteligenței artificiale care se ocupă cu construirea de programe care se îmbunătățesc automat pe baza experiențelor [81]. Direcția de cercetare pe care ne axăm în această teză este aplicarea modelelor dinamice de instruire automată pentru rezolvarea problemelor de clasificare supervizată și nesupervizată. Problemele concrete pe care am hotărât să le abordăm sunt detecția de pietoni (o problemă de clasificare supervizată) și clusterizarea expresiilor genice (o problemă de clasificare nesupervizată). Problemele abordate sunt reprezentative pentru cele două tipuri principale de clasificare și sunt foarte provocatoare, având o importanță foarte mare în viața reală.

Problema pe care o abordăm în cadrul direcției de cercetare a clasificării nesupervizate este clusterizarea dinamică a expresiilor genice. În zilele noastre sunt folosite tehnologii moderne pentru a măsura nivelele de expresie a mii de gene, în diferite condiții și de-a lungul timpului. O dată ce informațiile referitoare la expresiile genice sunt adunate, următorul pas este de a le analiza și de a extrage informații biologice utile, ceea ce se poate realiza prin intermediul clusterizării. În cazurilor seturilor de date cuprinzând expresii genice, fiecare genă este reprezentată de valorile expresiilor sale genice (attribute), la diferite momente de timp, în condițiile monitorizate.

În modele pe care le propunem, termenul *dinamic* se referă la faptul că setul de date nu este static, ci se modifică. Spre deosebire de abordările din literatură în care noi gene (instanțe) sunt adăugate în setul de date, noi tratăm cazul în care noi attribute (expresii genice pentru noi momente de timp) se adaugă la genele deja existente din setul de date. Din câte stim, nu exista abordări în literatură care tratează problema clusterizării dinamice de expresii genice, definită ca mai sus.

Problema pe care o abordăm în cadrul direcției de cercetare a clasificării supervizate este dezvoltarea de sisteme dinamice de recunoaștere a pietonilor. Siguranța pietonilor este o problemă majoră cu impact global, accidentele din trafic fiind una dintre cele mai mari cauze a deceselor și vătămărilor din întreaga lume. Pentru a proteja pietonii și biciclistii au fost dezvoltate sisteme avansate de asistare a conducătorului auto (ADAS). Acestea asistă conducătorul în luarea deciziilor, emit semnale în situații de conducere potențial periculoase și execută măsuri de contracarare. Pentru funcționarea bună a acestor sisteme ADAS este nevoie de sisteme eficiente de recunoaștere a pietonilor.

Pentru a dezvolta sisteme dinamice de recunoaștere a pietonilor care se adaptează condițiilor variabile de mediu, integrăm în modelele noastre dinamice informații din imagini multi-modale precum

intensitate, disparitate și flux optic. Se știe că prima necesitate a unui sistem de recunoaștere de pietoni este găsirea unui set robust de caracteristici extrase din imagini, care permite discriminarea precisă a formei umane, chiar și în condiții dificile. Ținând cont de acest aspect, dorim să investigăm o nouă tehnică pentru extragerea caracteristicilor din imagini, folosind kernel descriptori (KDs) [18], care au obținut rezultate bune în probleme de recunoaștere vizuală, dar din câte știm nu au fost folosite pentru detecția de pietoni.

Teza este organizată în patru capitole, după cum urmează:

În **Capitolul 1** prezentăm informațiile de bază pentru problema clasificării nesupervizate abordate în teză, problema clusterizării dinamice de expresii genice. Începem cu o privire de ansamblu asupra celor mai importante tehnici de clusterizare, cu accent pe algoritmi *k-means*, *fuzzy c-means* și *clusterizare ierarhică* pe care îi folosim în modele pe care le propunem. Continuăm apoi cu un tip special de clusterizare, clusterizarea dinamică, împreună cu o analiză a literaturii de specialitate din această direcție. În final, abordăm problema clusterizării dinamice de expresii genice și prezentăm o analiză a literaturii de specialitate.

Capitolul 2 este original și prezintă rezultatele cercetării noastre referitoare la problema clusterizării dinamice de expresii genice. Începem cu definirea problemei abordate și importanța acesteia în viața reală, apoi introducem trei algoritmi dinamici de clusterizare care pot gestiona expresiile genice nou colectate fără a relua execuția algoritmului de la început, pornind de la o partiție anterior obținută. În același context dinamic, propunem de asemenea un algoritm adaptiv pentru reguli de asociere relaționale pentru analiza expresiilor genice. Modele propuse sunt evaluate experimental pe un set de expresii genice, iar rezultatele sunt analizate și comparate cu rezultatele altor abordări dinamice din literatură. Rezultatele obținute subliniază eficiența modelelor noastre dinamice.

În **Capitolul 3** prezentăm informațiile de bază pentru problema clasificării supervizate abordate în teză, problema recunoașterii dinamice de pietoni. Începem prin a prezenta cele mai importante componente ale unui sistem de recunoaștere de pietoni: componenta de extragere de caracteristici și cea de clasificare. Prin urmare, vom prezenta succint cele mai frecvent utilizate caracteristici și clasificatori pentru recunoașterea de pietoni. În final, prezentăm o analiză a literaturii de specialitate în domeniu, cu accent pe fuziunea de modalități ale imaginilor, pentru care introducem în capitolul următor o abordare dinamică.

Capitolul 4 este original și prezintă rezultatele cercetării noastre referitoare la problema recunoașterii de pietoni în imagini uni și multi-modale. Începem cu o comparație a celor mai performante caracteristici în spectrul infra roșu, continuăm apoi cu o analiză a literaturii de specialitate cu privire la kernel descriptori, apoi prezentăm studiile noastre referitoare la utilizarea acestor caracteristici ca mod de reprezentare a imaginilor pentru recunoașterea de pietoni.

În primul studiu investigăm cum rezolvă problema recunoașterii de pietoni doi algoritmi de învățare, Mașinile cu Suport Vectorial (MSV) și Programarea Genetica (PG) folosind kerneli descriptori cu trei tipuri de kernele: exponențial, gaussian și laplacian, iar în al doilea studiu evaluăm performanța kernel descriptorilor în recunoașterea de pietoni uni și multi-modala. Propunem mai departe două modele dinamice pentru recunoașterea de pietoni care sunt capabile să selecteze cele mai discriminative modalități pentru fiecare imagine în parte, apoi să le includă în procesul de clasificare. Evaluările experimentale pe un set de imagini cu pietoni confirmă performanța modelelor noastre dinamice.

Contribuțiile originale din această teză sunt prezentate în Capitolele 2 and 4 și sunt următoarele:

- Trei algoritmi dinamici, care pot gestiona expresiile genice nou colectate pornind de la o partiție anterior obținută, fără a fi nevoie de a executa din nou algoritmul de la început.
 - Un algoritm bazat pe nuclee, care folosește clusterizarea *k-means* [22]; o euristică pentru determinarea numărului optim de clusteri într-un set de date cu expresii genice [22], evaluări experimentale ale algoritmului dinamic bazat pe nuclee pe un set de expresii genice reale, analiza rezultatelor și comparații cu rezultate obținute de alte modele dinamice din literatură [22].
 - Un algoritm dinamic pentru clusterizarea ierarhică a expresiilor genice, bazat pe algoritmul de clusterizare ierarhică [110]; evaluări experimentale ale algoritmului de clusterizare ierarhică dinamică pe un set de expresii genice reale, analiza rezultatelor și comparații cu rezultate obținute de alte modele dinamice din literatură [110].
 - Un algoritm dinamic pentru clusterizarea fuzzy a expresiilor genice, bazată pe algoritmul *fuzzy c-means* [111]; evaluări experimentale ale algoritmului de clusterizare dinamică fuzzy pe un set de expresii genice reale, analiza rezultatelor și comparații cu rezultate obținute de alte modele dinamice din literatură [111]
- Un algoritm dinamic pentru extragerea de reguli relaționale de asociere din expresii genice, care poate gestiona atributele nou obținute prin adaptarea setului de reguli obținut anterior, fără a fi nevoie de a relua execuția algoritmului de la început [31]; evaluări experimentale ale algoritmului de extragere a regulilor de asociere relaționale pe un set de expresii genice reale și analiza rezultatelor [31]
- Folosirea kernel descriptorilor pentru recunoașterea de pietoni [6, 112]
 - O comparație a modului în care doi algoritmi de instruirea automată: MSV și PG învață pe baza caracteristicilor extrase cu KD-uri folosind trei kenele: exponențial, gaussian și laplacian [6]
 - O comparație a performanței KD-urilor în imagini uni versus multi-modale, cu parametri optimizați independent pe fiecare modalitate în parte: intensitate, disparitate și flux optic [112]
- Doi algoritmi dinamici bazați pe instruirea automată, capabili de a determina dinamic modalitățile cele mai discriminative pentru clasificarea unei imagini
 - Un algoritm dinamic de selecție a modalităților, care reține cea mai discriminativă modalitate dintre intensitate, disparitate și flux optic [99]; evaluări experimentale ale algoritmului pe un set de date cu pietoni, analiza rezultatelor și comparații cu alte abordări din literatură [99].
 - Un algoritm dinamic de fuziune a modalităților, care fuzionează cele mai discriminative modalități dintre intensitate, disparitate și flux optic [99]; evaluări experimentale ale algoritmului pe un set de date cu pietoni, analiza rezultatelor și comparații cu alte abordări din literatură [113]

Concluzii

În această teză ne-am axat asupra aplicării modelelor dinamice de învățare supervizată pentru rezolvarea problemelor de clasificare supervizată și nesupervizată. Problemele concrete care am decis să le abordăm sunt clusterizarea expresiilor genice și recunoașterea de pietoni, deoarece sunt foarte provocatoare, au o importanță foarte mare în viața reală și sunt reprezentative pentru cele două tipuri principale de clasificare: supervizată și nesupervizată.

În prima direcție de cercetare am introdus trei algoritmi dinamici pentru clusterizarea expresiilor genice, în contextul în care expresii genice pentru noi momente de timp sunt adăugate la genele existente. Algoritmii (CBDCGE, DHCGE și FDCGE) sunt capabili de a adapta partiția anterior obținută când noi expresii genice sunt adăugate în setul de date, fără a reclusteriza de la început. Evaluările experimentale realizate pe un set de expresii genice reale au arătat că, în majoritatea cazurilor, clusterizarea este realizată mai eficient și de asemenea mai acurat folosind modelele noastre dinamice decât aplicând k-means, clusterizarea ierarhică, respectiv fuzzy c-means de la început pe setul de date extins. Cu toate acestea, există unele situații în care partițiile sunt prea dificil de adaptat după adăugarea de noi attribute și este recomandată o clusterizare completă.

În același context al expresiilor genice dinamice, am propus un algoritm adaptiv pentru extragerea de reguli de asociere relaționale (ARARM), care este capabil de a adapta setul de reguli descoperit anterior, în momentul în care noi expresii genice sunt adăugate în setul de date, fără a efectua de la zero extragerea de reguli. Experimente pe același set de expresii genice arată că ARARM ajunge mai rapid la soluție decât algoritmul de extragere de reguli executat de la început.

Direcții viitoare de cercetare vor fi în zona determinării unor condiții care să decidă când este mai potrivit să adaptezi (folosind CBDCGE, DHCGE sau FDCGE) partiția obiectelor cu attribute extinse decât să recalculezi partiția de la început, folosind un algoritm clasic de clusterizare. De asemenea dorim să extindem evaluarea experimentală pe alte seturi de date și să investigăm metode de a identifica automat distanța prag pentru clusteri (ex. folosind învățarea supervizată).

În a doua direcție de cercetare am abordat două probleme. Prima dată am realizat un studiu asupra eficienței folosirii kernel descriptorilor în recunoașterea de pietoni, datorită faptului că au obținut rezultate bune în recunoașterea automată și din câte știm nu au mai fost folosite pentru această problemă până acum. Apoi am introdus doi algoritmi dinamici, DMS și DMF, care sunt capabili de a determina cele mai discriminative modalități dintre intensitate, disparitate și flux optic pentru o imagine reprezentând o regiune de interes dintr-un cadru și mai departe să le includa în procesul de clasificare.

Kernel descriptorii au dovedit performanțe bune atât în imaginile uni cât și multi modale. Am optimizat parametrii kernelului pentru gradient independent pentru fiecare modalitate în parte, folosind

un algoritm în cascadă. Selecția celui mai potrivit kernel (exponential, gaussian și lapacian) depinde nu doar de imagini, ci și de algoritmul de învățare folosit. Chiar dacă KD-urile sunt considerate a fi o generalizare a histogramei orientării gradientilor, nu ating performanțele celor din urmă, fapt care poate fi cauzat de componenta de reducere a dimensionalității care folosește componenta în analiza principală pentru kernele (KPCA).

Evaluările experimentale pe un set de date cu pietoni arată ca modelele dinamice de selecție și fuziune, DMS și DMF sunt abordări promițătoare pentru recunoașterea de pietoni multi-modală. Primul are avantajele complexității reduse și a antrenamentului individual pe modalități, în timp ce al doilea obține performanțe mai bune. Mai mult, schemele de fuziune dinamice propuse în modele noastre sunt generice și pot fi aplicate în alte probleme care necesită o integrare dinamică a surselor.

Direcții viitoare de cercetare vor fi în zona extinderii și îmbunătățirii modelelor dinamice propuse (ex. integrând indicatori de relevanță a modalităților bazați pe imagini, adăugarea altor caracteristici în fuziune sau folosirea altora) și de asemenea evaluării acestora pe alte seturi de date. Mai departe pot fi aduse îmbunătățiri kernel descriptorilor prin optimizarea kernelelor gaussiene (de poziție și de orientare din cadrul kernelului de gradient) și adaptarea procesului de reducere a dimensionalității cu scopul de a reține cele mai relevante informații.

Bibliografie

- [1] Bryant Aaron, Dan E. Tamir, Naphtali D. Rishé, and Abraham Kandel. Dynamic incremental fuzzy c-means clustering. In *The Sixth International Conferences on Pervasive Patterns and Applications*, 2014.
- [2] Bryant Aaron, Dan E. Tamir, Naphtali D. Rishé, and Abraham Kandel. Dynamic incremental k-means clustering. In *International Conference on Computational Science and Computational Intelligence*, 2014.
- [3] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, dec 2006.
- [4] S. Albayrak and F. Amasyali. Fuzzy c-means clustering on medical diagnostic systems. In *Turkish Symposium on Artificial Intelligence and Neural Networks - TAINN*, 2003.
- [5] L. An and R.W. Doerge. Dynamic clustering of gene expression. *ISRN Bioinformatics*, 2012:1–12, 2012.
- [6] A. Andreica, L. Diosan, R. D. Gaceanu, and A. Sirbu. Pedestrian recognition by using kernel descriptors. *Studia Universitatis Babeş-Bolyai, Seria Informatica*, LVIII(2):77–89, 2013.
- [7] Anelia Angelova, Alex Krizhevsky, and Vincent Vanhoucke. Pedestrian detection with a large-field-of-view deep network. In *ICRA*, pages 704–711. IEEE, 2015.
- [8] A. Apatean, C. Rusu, A. Rogozan, and A. Bensrhair. Visible-infrared fusion in the frame of an obstacle recognition system. In *Automation Quality and Testing Robotics (AQTR), Cluj-Napoca*, pages 1 – 6, 2010.
- [9] C. Arima, T. Hanai, and M. Okamoto. Gene Expression Analysis Using Fuzzy K-Means Clustering. *Genome Informatics*, 14:334–335, 2003.
- [10] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, and et al. Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nature Genetics*, 25(1):25–29, 2000.
- [11] Abraham Bagherjeiran, Christoph F. Eick, Chun-Sheng Chen, and Ricardo Vilalta. Adaptive clustering: Obtaining better clusters using feedback and past experience. In *Proceedings of the Fifth IEEE International Conference on Data Mining, ICDM '05*, pages 565–568, Washington, DC, USA, 2005. IEEE Computer Society.

- [12] A.M. Bagirov and K. Mardaneh. Modified global k-means algorithm for clustering in gene expression data sets. In *Proceedings of the 2006 workshop on Intelligent systems for bioinformatics, WISB '06*, pages 23–28, Darlinghurst, Australia, Australia, 2006. Australian Computer Society, Inc.
- [13] Z. Bar-Joseph, G. Gerber, D.K. Gifford, and T.S. Jaakkola. A New Approach to Analyzing Gene Expression Time Series Data. In *Proceedings of the sixth annual international conference on Computational biology, RECOMB '02*, pages 39–48, 2002.
- [14] H. Bay, T. Tuytelaars, and L. J. Van Gool. Surf speeded up robust features. In *ECCV*, pages 404–417, 2006.
- [15] M. Bertozzi, A. Broggi, M. Felisa, and G. Vezzoni. Low-level pedestrian detection by means of visible and far infra-red tetra-vision. *IEEE Intelligent Vehicles Symposium*, pages 231–236, 2006.
- [16] Bassem Besbes, Sonda Ammar, Yousri Kessentini, Alexandrina Rogozan, and Abdelaziz Benschair. Evidential combination of SVM road obstacle classifiers in visible and far infrared images. In *Intelligent Vehicles Symposium*, pages 1074–1079. IEEE, 2011.
- [17] Bassem Besbes, Alexandrina Rogozan, Adela-Maria Rus, Abdelaziz Benschair, and Alberto Broggi. Pedestrian detection in far-infrared daytime images using a hierarchical codebook of surf. *Sensors*, 15(4):8570, 2015.
- [18] Liefeng Bo, Xiaofeng Ren, and Dieter Fox. Kernel descriptors for visual recognition. In *NIPS*, pages 244–252. Curran Associates, Inc, 2010.
- [19] Liefeng Bo, Xiaofeng Ren, and Dieter Fox. Depth kernel descriptors for object recognition. In *IROS*, pages 821–826. IEEE, 2011.
- [20] Liefeng Bo, Xiaofeng Ren, and Dieter Fox. Object recognition with hierarchical kernel descriptors. In *CVPR*, pages 1729–1736, 2011.
- [21] Liefeng Bo and Cristian Sminchisescu. Efficient match kernel between sets of features for visual recognition. In Yoshua Bengio, Dale Schuurmans, John D. Lafferty, Christopher K. I. Williams, and Aron Culotta, editors, *NIPS*, pages 135–143. Curran Associates, Inc, 2009.
- [22] Maria Iuliana Bocicor, Adela Sirbu, and Gabriela Czibula. Dynamic core based clustering of gene expression data. *International Journal of Innovative Computing, Information and Control*, 10(3):1051–1069, 2014.
- [23] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In D. Haussler, editor, *COLT*, pages 144–152. ACM Press, 1992.
- [24] LD Brown, TT Cat, and A DasGupta. Interval estimation for a proportion. *Statistical Science*, 16:101–133, 2001.
- [25] Toon Calders, Nele Dexters, Joris J. M. Gillis, and Bart Goethals. Mining frequent itemsets in a stream. *Inf. Syst*, 39:233–255, 2014.

- [26] Alina Campan, Gabriela Serban, and Adrian Marcus. Relational association rules and error detection. *Informatica*, LI(1):31–36, 2006.
- [27] Alina Campan, Gabriela Serban, Traian Marius Truta, and Andrian Marcus. An algorithm for the discovery of arbitrary length ordinal association rules. In *DMIN*, pages 107–113, 2006.
- [28] C. C. Chang and C. J. Lin. *LIBSVM: a library for support vector machines*. Online, 2001.
- [29] Moses Charikar, Chandra Chekuri, Tomas Feder, and Raajev Motwani. Incremental clustering and dynamic information retrieval. *SICOMP: SIAM Journal on Computing*, 33, 2004.
- [30] Gabriela Czibula, Maria-Iuliana Bocicor, and Istvan Gergely Czibula. Promoter sequences prediction using relational association rule mining. *Evolutionary Bioinformatics*, 8:181–196, 04 2012.
- [31] Gabriela Czibula, Istvan-Gergely Czibula, Adela Sirbu, and Gabriel Mircea. A novel approach to adaptive relational association rule mining. *Applied Soft Computing*, 36:519–533, November 2015.
- [32] Gabriela Czibula, Zsuzsanna Marian, and István Gergely Czibula. Detecting software design defects using relational association rule mining. *Knowl. Inf. Syst*, 42(3):545–577, 2015.
- [33] Gabriela Czibula, Zsuzsanna Marian, and István Gergely Czibula. Detecting software design defects using relational association rule mining. *Knowl. Inf. Syst*, 42(3):545–577, 2015.
- [34] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, pages I: 886–893, 2005.
- [35] N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. In *ECCV*, pages II: 428–441, 2006.
- [36] R. Das, D.K. Bhattacharyya, and J.K. Kalita. An Incremental Clustering of Gene Expression data. *World Congress on Nature and Biologically Inspired Computing. NaBIC 2009.*, pages 742–747, 2009.
- [37] R. Das, J.K. Kalita, and D.K. Bhattacharyya. A new approach for clustering gene expression time series data. *International Journal of Bioinformatics Research and Applications*, 5(3):310–328, 2009.
- [38] Rosy Das, Jugal K. Kalita, and Dhruva Kumar Bhattacharyya. A pattern matching approach for clustering gene expression data. *IJDMMM*, 3(2), 2011.
- [39] Ian Davidson, S. S. Ravi, and Martin Ester. Efficient incremental constrained clustering. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '07, pages 240–249, New York, NY, USA, 2007. ACM.
- [40] Li Deng and Dong Yu. Deep learning: Methods and applications. *Foundations and Trends in Signal Processing*, 7(3-4):197–387, 2014.

- [41] J.L. DeRisi, P.O. Iyer, and V.R. Brown. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, 278(5338):680–686, 1997.
- [42] Laura Diosan, Alexandrina Rogozan, and Jean-Pierre Pecuchet. Improving classification performance of support vector machine by genetically optimising kernel shape and hyper-parameters. *Appl. Intell*, 36(2):280–294, 2012.
- [43] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):743–761, 2012.
- [44] Markus Enzweiler, Angela Eigenstetter, Bernt Schiele, and Dariu M. Gavrilă. Multi-cue pedestrian classification with partial occlusion handling. In *CVPR*, pages 990–997. IEEE, 2010.
- [45] Markus Enzweiler and Dariu M. Gavrilă. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2179–2195, 2009.
- [46] Markus Enzweiler and Dariu M. Gavrilă. A multilevel mixture-of-experts framework for pedestrian classification. *IEEE Transactions on Image Processing*, 20(10):2967–2979, 2011.
- [47] Markus Enzweiler, Pascal Kanter, and Dariu M. Gavrilă. Monocular Pedestrian Recognition Using Motion Parallax. In *Proc. IEEE Symposium on Intelligent Vehicles*, pages 792–797, 2008.
- [48] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. Liblinear: A library for large linear classification. *Journal of Machine Learning Research*, (9):1871–1874, 2008.
- [49] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, June 2006.
- [50] H.-J. Frasch, M.H. Medema, E. Takano, and R. Breitling. Design-based re-engineering of biosynthetic gene clusters: plug-and-play in practice. *Current Opinion in Biotechnology*, 2013.
- [51] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *JCSS: Journal of Computer and System Sciences*, 55, 1997.
- [52] D. M. Gavrilă and S. Munder. Multi-cue pedestrian detection and tracking from a moving vehicle. *International Journal of Computer Vision*, 73(1):41–59, jun 2007.
- [53] David Gerónimo, Antonio M. López, Angel Domingo Sappa, and Thorsten Graf. Survey of pedestrian detection for advanced driver assistance systems. *IEEE Trans. Pattern Anal. Mach. Intell*, 32(7):1239–1258, 2010.
- [54] V. Gesu, R. Giancarlo, G. Lo Bosco, A. Raimondi, and D. Scaturro. Genclust: A genetic algorithm for clustering gene expression data. *BMC Bioinformatics*, 6(289), 2005.
- [55] F.D. Gibbons and F.P. Roth. Judging the quality of gene expression-based clustering methods using gene annotation. *Genome Research*, 12(10):1574–1581, 2002.

- [56] D. Goel and T. H. Chen. Pedestrian detection using global-local motion patterns. In *ACCV*, pages I: 220–229, 2007.
- [57] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*, volume 54. Morgan Kaufmann, 2006.
- [58] R. Henson and L. Cetto. *The MATLAB bioinformatics toolbox. Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*. The MathWorks Inc., Natick, Massachusetts, 2005.
- [59] J. Herrero and J. Dopazo. Combining hierarchical clustering and self-organizing maps for exploratory analysis of gene expression patterns. *Journal of Proteome Research*, 1(5):467–470, 2002.
- [60] H. Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *CVPR*, pages II: 807–814, 2005.
- [61] Di Huang, Caifeng Shan, Mohsen Ardabilian, Yunhong Wang, and Liming Chen. Local binary patterns and its application to facial image analysis: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 41(6), 2011.
- [62] A. K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recogn. Lett.*, 31(8):651–666, 2010.
- [63] Anil K. Jain and Richard C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, 1988.
- [64] J.-Y. Jiang, W.-H. Cheng, and S.-J. Lee. A dissimilarity measure for document clustering. *ICIC Express Letters*, 6(1):15–21, 2012.
- [65] L. P. Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285, 04 1996.
- [66] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics, volume III*. Griffin, London, 1966.
- [67] K. Kim, S. Zhang, K. Jiang, L. Cai, I.B. Lee, L.J. Feldman, and H. Huang. Measuring similarities between gene expression profiles through new data transformations. *BMC Bioinformatics*, 8(29), 2007.
- [68] Dana Kulic, Wataru Takano, and Yoshihiko Nakamura. Combining automated on-line segmentation and incremental clustering for whole body motions. In *ICRA*, pages 2591–2598. IEEE, 2008.
- [69] Christoph H. Lampert. Kernel methods in computer vision. *Foundations and Trends in Computer Graphics and Vision*, 4(3):193–285, 2009.
- [70] G. N. Lance and W. T. Williams. A general theory of classificatory sorting strategies II. Clustering systems. *The Computer Journal*, 10(3):271–277, November 1967.

- [71] Taoying Li and Yan Chen. Fuzzy K-means incremental clustering based on K-center and vector quantization. *JCP*, 5(11):1670–1677, 2010.
- [72] Wei Li, Rujing Wang, Liangtu Song, and Xiufang Jia. Batch dynamically incremental c-means clustering algorithm based on rough fuzzy set. *Journal of Computational Information Systems*, 11(5):1553–1561, 2015.
- [73] Yan Li, Snigdha Verma, Li Lao, and Jun-Hong Cui. SACA: SCM-based adaptive clustering algorithm. In *MASCOTS*, pages 271–279. IEEE Computer Society, 2005.
- [74] Dimitri A. Lisin, Marwan A. Mattar, Matthew B. Blaschko, Mark C. Benfield, and Erik G. Learned-Miller. Combining local and global image features for object class recognition. In *In Proceedings of the IEEE CVPR Workshop on Learning in Computer Vision and Pattern Recognition*, pages 47–55, 2005.
- [75] L. Lopez-Kleine, J. Romeo, and F. Torres-Avils. Gene functional prediction using clustering methods for the analysis of tomato microarray data. In *7th International Conference on Practical Applications of Computational Biology and Bioinformatics Advances in Intelligent Systems and Computing*, volume 222, pages 1–6, 2013.
- [76] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [77] Y. Lu, S. Lu, F. Fotouhi, Y. Deng, and S.J. Brown. Incremental genetic k-means algorithm and its application in gene expression data analysis. *BMC Bioinformatics*, 5(172), 2004.
- [78] Y. Luan and H. Li. Clustering of time-course gene expression data using a mixed-effects model with b-splines. *Bioinformatics*, 19(4):474–482, 2003.
- [79] P. C. Mahalanobis. On the generalised distance in statistics. In *Proc. of the Nat. Inst. of Sci. of India*, volume 2, pages 49–55, 1936.
- [80] T. Mitchell. *Machine Learning (Mcgraw-Hill International Edit)*. McGraw-Hill Education (ISE Editions), 1st edition, October 1997.
- [81] Tom Mitchell. *Machine Learning*. McGraw-Hill, 1997.
- [82] S. Munder and S. Gavrilu. An experimental study on pedestrian classification. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(11):1863–1868, nov 2006.
- [83] S. Munder, C. Schnorr, and D. M. Gavrilu. Pedestrian detection and tracking using a mixture of view-based shape-texture models. *IEEE Trans. Intelligent Transportation Systems*, 9(2):333–343, June 2008.
- [84] S. Nedeveschi, S. Bota, and C. Tomiuuc. Stereo-based pedestrian detection for collision-avoidance applications. *IEEE Trans. Intelligent Transportation Systems*, 10(3):380–391, sep 2009.
- [85] T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59, jan 1996.

- [86] K. Okuma, A. Taleghani, N. de Freitas, J. J. Little, and D. G. Lowe. A boosted particle filter: Multitarget detection and tracking. In *ECCV*, pages Vol I: 28–39, 2004.
- [87] Luciano Oliveira, Urbano Nunes, and Paulo Peixoto. On exploration of classifier ensemble synergism in pedestrian detection. *IEEE Transactions on Intelligent Transportation Systems*, 11(1):16–27, 2010.
- [88] David L. Olson and Dursun Delen. *Advanced Data Mining Techniques*. Springer, 2008.
- [89] Mihai Oltean. Improving the search by encoding multiple solutions in a chromosome. In Nadia Nedjah and Luiza de Macedo Mourelle, editors, *Evolutionary Machine Design: Methodology and Applications*, Intelligent System Engineering, chapter 4, pages 85–110. Nova Publishers, 2005.
- [90] Mihai Oltean, Crina Grosan, Laura Diosan, and Cristina Mihaila. Genetic programming with linear representation: a survey. *International Journal on Artificial Intelligence Tools*, 18(2):197–238, 2009.
- [91] M. Oren, C. P. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio. Pedestrian detection using wavelet templates. In *CVPR*, pages 193–199, 1997.
- [92] Wanli Ouyang and Xiaogang Wang. Joint deep learning for pedestrian detection. In *ICCV*, pages 2056–2063. IEEE, 2013.
- [93] S. Paisitkriangkrai, C. H. Shen, and J. Zhang. Fast pedestrian detection using a cascade of boosted covariance features. *IEEE Trans. Circuits and Systems for Video Technology*, 18(8):1140–1151, August 2008.
- [94] M. K. Pakhira, S. Bandyopadhyay, and U. Maulik. Validity index for crisp and fuzzy clusters. *Pattern Recognition*, 37(3):487 – 501, 2004.
- [95] I. Parra Alonso, D. Fernandez Llorca, M. A. Sotelo, L. M. Bergasa, P. Revenga de Toro, J. Nuevo, M. Ocana, and M. A. Garcia Garrido. Combination of feature extraction methods for SVM pedestrian detection. *IEEE Trans. Intelligent Transportation Systems*, 8(2):292–307, April 2007.
- [96] V.R. Pejaver, H. Lee, and S. Kim. Gene cluster prediction and its application to genome annotation. *Protein Function Prediction for Omics Era*, pages 35–54, 2011.
- [97] J. Platt. Fast training of support vector machines using sequential minimal optimization. In *Advances in Kernel Methods — Support Vector Learning*, pages 185–208, Cambridge, MA, 1999. MIT Press.
- [98] Marcus Rohrbach, Markus Enzweiler, and Dariu M. Gavrilă. High-level fusion of depth and intensity for pedestrian classification. In *DAGM-Symposium*, volume 5748 of *Lecture Notes in Computer Science*, pages 101–110. Springer, 2009.
- [99] Adela-Maria Rus, Alexandrina Rogozan, Laura Diosan, and Abdelaziz Bensrhair. Pedestrian recognition by using a dynamic modality selection approach. In *IEEE 18th International Conference on Intelligent Transportation Systems, Gran Canaria, Spain, September 15-18*, pages 1862 – 1867, 2015.

- [100] S. Sarmah and D.K. Bhattacharyya. An effective technique for clustering incremental gene expression data. *International Journal of Computer Science Issues*, 7(3):31–41, 2010.
- [101] Robert E. Schapire. Explaining adaboost, 08 2013.
- [102] B. Schölkopf. The kernel trick for distances. In *NIPS*, pages 301–307, Cambridge, MA, 2000. MIT Press.
- [103] B. Scholkopf, A. Smola, and K. Muller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, (10):1299–1319, 1998.
- [104] G. Serban and A. Campan. Incremental clustering using a core-based approach. In *Proceedings of the 20th international conference on Computer and Information Sciences, ISCIS'05*, pages 854–863, Berlin, Heidelberg, 2005. Springer-Verlag.
- [105] Gabriela Serban and Alina Campan. Hierarchical adaptive clustering. *Informatica*, 19(1):101–112, 2006.
- [106] Gabriela Serban, Alina Campan, and Istvan Gergely Czibula. A programming interface for finding relational association rules. *International Journal of Computers, Communications and Control*, I(S.):439–444, jun 2006.
- [107] Pierre Sermanet, Koray Kavukcuoglu, and Yann Lecun. Traffic signs and pedestrians vision with multi-scale convolutional networks. In *In Snowbird Machine Learning Workshop*, 2011.
- [108] Archana Singh, Avantika Yadav, and Ajay Rana. K-means with three different distance metrics. *International Journal of Computer Applications*, 67(10):13–17, 2013.
- [109] Adela Sirbu. A study on dynamic clustering of gene expression data. *Informatica*, LIX(1):16–27, 2014.
- [110] Adela Sirbu and Maria-Iuliana Bocicor. A dynamic approach for hierarchical clustering of gene expression data. In *Intelligent Computer Communication and Processing (ICCP)*, pages 3–6, Sept 2013.
- [111] Adela-Maria Sirbu, Gabriela Czibula, and Maria-Iuliana Bocicor. Dynamic clustering of gene expression data using a fuzzy approach. In *16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2014, Timisoara, Romania, September 22-25, 2014*, pages 220–227. IEEE, 2014.
- [112] Adela-Maria Sirbu, Alexandrina Rogozan, Laura Diosan, and Abdelaziz Bensrhair. Pedestrian recognition by using a kernel-based multi-modality approach. In *16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2014, Timisoara, Romania, September 22-25, 2014*, pages 258–263. IEEE, 2014.
- [113] Adela-Maria Sirbu, Alexandrina Rogozan, Laura Diosan, and Abdelaziz Bensrhair. Pedestrian recognition using a dynamic modality fusion approach. In *Proceedings of IEEE International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, September 3-5*, pages 393 – 400, 2015.

- [114] B. Song and H. Lee. Prioritizing disease genes by integrating domain interactions and disease mutations in a protein-protein interaction network. *International Journal of Innovative Computing, Information and Control*, 8(2):1327–1338, 2012.
- [115] D. Stekel. *Microarray Bioinformatics*. Cambridge University Press, Cambridge, UK, 2006.
- [116] Xiaoke Su, Yang Lan, Renxia Wan, and Yuming Qin. A fast incremental clustering algorithm. In *Proceedings of the 2009 International Symposium on Information Processing*, pages 175–178, 2009.
- [117] P. Tamayo, D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E. Lander, and T. Golub. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proceedings of National Academy of Sciences*, 96(6):2907–2912, 1999.
- [118] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining*. Addison-Wesley, 2005.
- [119] Stephane Tuffery. *Data Mining and Statistics for Decision Making*. John Wiley and Sons, 2011.
- [120] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.
- [121] V. Vapnik. *Statistical Learning Theory*. Wiley, 1998.
- [122] Renato Vimieiro and Pablo Moscato. A new method for mining disjunctive emerging patterns in high-dimensional datasets using hypergraphs. *Inf. Syst*, 40:1–10, 2014.
- [123] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Proc. CVPR*, 1:511–518, 2001.
- [124] P. Viola, M. J. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. *International Journal of Computer Vision*, 63(2):153–161, jul 2005.
- [125] P. Viola and M.J. Jones. Robust real-time face detection. *International Journal of Computer Vision (IJCV)*, 57(2):137–154, 2004.
- [126] Paul Viola and Michael Jones. Robust real-time object detection. In *International Journal of Computer Vision*, 2001.
- [127] Stefan Walk, Nikodem Majer, Konrad Schindler, and Bernt Schiele. New features and insights for pedestrian detection. pages 1030–1037. IEEE Computer Society, 2010.
- [128] Stefan Walk, Konrad Schindler, and Bernt Schiele. Disparity statistics for pedestrian detection: Combining appearance, motion and stereo. In *ECCV (6)*, volume 6316, pages 182–195. Springer, 2010.
- [129] Peng Wang, Jingdong Wang, Gang Zeng, Weiwei Xu, Hongbin Zha, and Shipeng Li. Supervised kernel descriptors for visual recognition. In *CVPR*, pages 2858–2865. IEEE, 2013.

- [130] Xiaoyu Wang, Tony X. Han, and Shuicheng Yan. An HOG-LBP human detector with partial occlusion handling. In *ICCV*, pages 32–39. IEEE, 2009.
- [131] Xiaoyu Wang, Tony X. Han, and Shuicheng Yan. An HOG-LBP human detector with partial occlusion handling. In *ICCV*, pages 32–39. IEEE, 2009.
- [132] Andreas Wedel, Daniel Cremers, Thomas Pock, and Horst Bischof. Structure- and motion-adaptive regularization for high accuracy optic flow. In *ICCV*, pages 1663–1668. IEEE, 2009.
- [133] WHO. Global status report on road safety. *World Health Organization*, 2015.
- [134] C. Wojek and B. Schiele. A performance evaluation of single and multi-feature people detection. *DAGM Symposium Pattern Recognition*, pages 82–91, 2008.
- [135] Fei Wu and Georges Gardarin. Gradual clustering algorithms. In *DASFAA*, pages 48–55. IEEE Computer Society, 2001.
- [136] X. Xiao, E.R. Dow, R.C. Eberhart, Z.B. Miled, and R.J. Oppelt. Gene Clustering Using Self-Organizing Maps and Particle Swarm Optimization. In: *Proc. 17th Intl. Symposium on Parallel and Distributed Processing*, 2003.
- [137] Guandong Xu, Yu Zong, and Zhenglu Yang. *Applied Data Mining*. CRC Press, Inc., Boca Raton, FL, USA, 2013.
- [138] J. Yang and J. Watada. Fuzzy clustering analysis of data mining: Application to an accident mining system. *International Journal of Innovative Computing, Information and Control*, 8(8):5715–5724, 2012.
- [139] N. Yano and M. Kotani. Clustering gene expression data using self-organizing maps and k-means clustering. *Proceedings of SICE 2003 Annual Conference*, 3:3211–3215, 2003.
- [140] Steven Young, Itamar Arel, Thomas P. Karnowski, and Derek Rose. A fast and stable incremental clustering algorithm. In Shahram Latifi, editor, *Seventh International Conference on Information Technology: New Generations, ITNG 2010, Las Vegas, Nevada, USA, 12-14 April 2010*, pages 204–209. IEEE Computer Society, 2010.
- [141] Y. Yuhui, C. Lihui, A. Goh, and A. Wong. Clustering gene data via associative clustering neural network. In: *Proc. 9th Intl. Conf. on Information Processing*, pages 2228–2232, 2002.
- [142] L. Zhao and C. E. Thorpe. Stereo- and neural network-based pedestrian detection. *IEEE Trans. Intelligent Transportation Systems*, 1(3):148–154, September 2000.
- [143] Liang Zhao and Charles E. Thorpe. Stereo- and neural network-based pedestrian detection. *IEEE Transactions on Intelligent Transportation Systems*, 1(3):148–154, 2000.