10302 Abstracts Collection
Learning paradigms in dynamic environments
— Dagstuhl Seminar —

Barbara Hammer[1], Pascal Hitzler[2], Wolfgang Maass[3] and Marc Toussaint[4]

[1] Universität Bielefeld, DE
bhammer@techfak.uni-bielefeld.de
[2] Wright State University, Dayton, US
pascal@pascal-hitzler.de,pascal.hitzler@wright.edu
[3] TU Graz, AT
[4] TU Berlin, DE
mtoussai@cs.tu-berlin.de

Abstract. From 25.07. to 30.07.2010, the Dagstuhl Seminar 10302 "Learning paradigms in dynamic environments " was held in Schloss Dagstuhl – Leibniz Center for Informatics. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general. Links to extended abstracts or full papers are provided, if available.

Keywords. Recurrent neural networks, Dynamic systems, Speech processing, Neurobiology, Neural-symbolic integration, Autonomous learning

10302 Summary – Learning paradigms in dynamic environments

The seminar centered around problems which arise in the context of machine learning in dynamic environments. Particular emphasis was put on a couple of specific questions in this context: how to represent and abstract knowledge appropriately to shape the problem of learning in a partially unknown and complex environment and how to combine statistical inference and abstract symbolic representations; how to infer from few data and how to deal with non i.i.d. data, model revision and life-long learning; how to come up with efficient strategies to control realistic environments for which exploration is costly, the dimensionality is high and data are sparse; how to deal with very large settings; and how to apply these models in challenging application areas such as robotics, computer vision, or the web.

Dagstuhl Seminar Proceedings 10302
Learning paradigms in dynamic environments
http://drops.dagstuhl.de/opus/volltexte/2010/2804

Keywords: Summary

Joint work of: Hammer, Barbara; Hitzler, Pascal; Maass, Wolfgang; Toussaint, Marc

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2010/2802

Hierarchical Models for Image Interpretation

Sven Behnke (Universität Bonn, DE)

In the talk, I introduce a hierarchical architecture for iterative image interpretation.

This architecture represents images at multiple levels of abstraction. The spatial resolution of the representations decreases with height while feature diversity and invariance increase. The representations are computed by a recurrent convolutional neural network. The network has many horizontal (between adjacent locations at the same level) and vertical (between the same location at adjacent levels) loops, which allow for the flexible incorporation of contextual information when resolving local ambiguities. An image is interpreted first at positions where little ambiguity exists. Partial results then bias the interpretation of the more ambiguous stimuli. Such a refinement is most useful when image contrast is low, noise and distractors are present, objects are partially occluded, or the interpretation is otherwise complicated.

The proposed architecture can be trained using unsupervised and supervised learning techniques. Applications include super-resolution, filling-in of occlusions, contrast enhancement and noise removal, face localization, and object categorization.

The architecture has been implemented on GPUs. The significant speedup makes it possible to learn large unconstrained object recognition tasks, like ImageNet.

Keywords: Computer vision, object recognition, neural networks

Neural-Symbolic Learning Systems: Neural Networks for Normative Reasoning

Silvano Colombo-Tosatto (University of Torino, IT)

We present how (a fragment of) Input/Output logic can be embedded into feed forward Neural Networks by using the neural-symbolic methodology. We do so in order to exploit the neural networks capabilities for normative reasoning. We show how a knowledge base of Input/Output logic rules can be represented in a neural network.

Finally we show how a neural network can be used to reason about dilemma and contrary to duty problems, common problems of normative reasoning.

Keywords: Neural-Symbolic Approach, Neural Networks, Input/Output Logic, Normative Reasoning

kLog — A Language for Logical and Relational Learning with Kernels

Paolo Frasconi (University of Firenze, IT)

KLog is a logical and relational language language for kernel-based learning. It builds on simple but powerful concepts: learning from interpretations, entity/relationship data modeling, logic programming and deductive databases (Prolog and Datalog), and graph kernels. kLog is a statistical relational learning system but unlike other statistical relational learning models, it does not represent a probability distribution. It is rather a kernel-based approach to learning that employs features derived from a grounded entity/relationship diagram. These features are derived using a novel technique called *graphicalization* that is used to transform the relational representations in a graph based representations. Once the graphs are computed, kLog employs graph kernels for learning. kLog can use numerical and symbolic data, background knowledge in the form of Prolog or Datalog programs (as in inductive logic programming systems) and several statistical procedures can be used to fit the model parameters. The kLog framework can be applied to tackle the same range of tasks that has made statistical relational learning so popular, including classification, regression, multi-task learning, and collective classification.

Joint work of: Frasconi, Paolo; Costa Fabrizio; De Raedt, Luc; De Grave, Kurt

Neurons and Symbols: A Manifesto

Artur Garcez (City University - London, GB)

We discuss the purpose of neural-symbolic integration including its principles, mechanisms and applications. We outline a cognitive computational model for neural-symbolic integration, position the model in the broader context of multi-agent systems, machine learning and automated reasoning, and list some of the challenges for the area of neural-symbolic computation to achieve the promise of effective integration of robust learning and expressive reasoning under uncertainty.

Keywords: Neuro-symbolic systems, cognitive models, machine learning

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2010/2800

Some steps towards a general principle for dimensionality reduction mappings

Barbara Hammer (Universität Bielefeld, DE)

In the past years, many dimensionality reduction methods have been established which allow to visualize high dimensional data sets. Recently, also formal evaluation schemes have been proposed for data visualization, which allow a quantitative evaluation along general principles. Most techniques provide a mapping of a priorly given finite set of points only, requiring additional steps for out-of-sample extensions. We propose a general view on dimensionality reduction based on the concept of cost functions, and, based on this general principle, extend dimensionality reduction to explicit mappings of the data manifold. This offers the possibility of simple out-of-sample extensions. Further, it opens a way towards a theory of data visualization taking the perspective of its generalization ability to new data points. We demonstrate the approach based in a simple example.

Keywords: Visualization, dimensionality reduction

Joint work of: Hammer, Barbara; Bunte, Kerstin; Biehl, Michael

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2010/2803

Some ideas about dimensionality reduction mappings

Barbara Hammer (Universität Bielefeld, DE)

Dimensionality reduction is closely connected to autonomous learning because of three reasons: it constitutes an indispensible tool to avoid the curse of dimensionality which is often present in autonomous learning tasks due to complex data and multimodalities, it offers an efficient and intuitive way for humans to inspect important aspects of an autonomous system such as an internal state, and it shares key challenges of autonomous learning, in particular an inherently ill-posed problem which needs appropriate shaping. Apart from these properties, however, a formal theory of dimensionality recution and, in particular, its formal evaluation and proofs of its suitability are so far rare.

In the talk, we review several popular dimensionality reduction methods and formalize the underlying principle as a principle of cost optimization with different foci. This opens the way towards a generalization of dimensionality reduction techniques, which often map a given finite set of data points only, to dimensionality reduction mappings, as well as a way to formalize and bound their generalization ability by means of statistical learning theory.

Learning for Reasoning - Learning as Reasoning - for the Semantic Web

Pascal Hitzler (Wright State University - Dayton, US)

The realization of Semantic Web reasoning is central to substantiating the Semantic Web vision. However, current mainstream research on this topic faces serious challenges, which forces us to question established lines of research and to rethink the underlying approaches. In particular, we argue that established deductive reasoning techniques for the Semantic Web need to be complemented by machine learning approaches. The conceptual rationale underlying this change of perspective lies in a recasting of reasoning problems into classification and information retrieval tasks.

Pascal Hitzler, Frank van Harmelen, A Reasonable Semantic Web. Semantic Web - Interoperability, Usability, Applicability. To appear. Preprint available from http://www.semantic-web-journal.net

Keywords: Semantic Web, Automated Reasoning, Machine Learning

Full Paper:
http://www.semantic-web-journal.net/content/new-submission-reasonable-semantic-web

See also: Pascal Hitzler, Frank van Harmelen, A Reasonable Semantic Web. Semantic Web - Interoperability, Usability, Applicability. To appear. Preprint available from http://www.semantic-web-journal.net

Second-order SMO for SVM Online Learning

Christian Igel (Ruhr-Universität Bochum, DE)

The LASVM by Bordes et al. (2005) iteratively approximates the SVM solution using sequential minimal optimization (SMO). It allows for online and active learning. In each SMO step, LASVM uses the most violating pair heuristic for choosing a pairs of variables for optimization. We propose to replace this strategy by a second order method, which greedily maximizes the progress in the dual in each single step. Our online algorithm learns faster and finds sparser hypotheses, because even a few unfavorable decisions by the first-order heuristic may impair the course of optimization.

Keywords: Support vector machine, online learning

Joint work of: Igel, Christian; Glasmachers, Tobias

Full Paper:
http://www.neuroinformatik.ruhr-uni-bochum.de/PEOPLE/igel/SOSMOISVMOaAL.pdf

See also: Tobias Glasmachers, Christian Igel: Second-Order SMO Improves SVM Online and Active Learning, Neural Computation 20(2): 374-382, 2008

Neuroevolution Strategies

Christian Igel (Ruhr-Universität Bochum, DE)

We propose neuroevolution strategies (NeuroESs) for direct policy search in reinforcement learning (RL). The covariance matrix adaptation evolution strategy (CMA-ES, Hansen and Ostermeier, Evolutionary Computation 9, 2001) is applied to adapt the weight of neural network policies. Learning using NeuroESs can be interpreted as modelling of extrinsic perturbations on the level of synaptic weights. In contrast, policy gradient methods can be regarded as intrinsic perturbation of neuronal activity.

Several (independent) evaluations of the CMA-ES for RL demonstrate excellent performance, in particular for episodic tasks without helpful intermediate rewards. The CMA-ES for RL (1) is a variable metric algorithm learning an appropriate coordinate system for searching better policies, (2) extracts a search direction from the scalar reward signals, (3) is comparatively robust w.r.t. tuning of meta-parameters, (4) is based on ranking policies, which is less susceptible to uncertainty & noise compared to estimating a value function or a gradient of a performance measure w.r.t. policy parameters, (5) is applicable if function approximators are non-differentiable, and (6) can be regarded as an efficient implementation of biological learning principles.

Keywords: Reinforcement learning, covariance matrix adaptation evolution strategy, direct policy search

Joint work of: Igel, Christian; Heidrich-Meisner, Verena

See also: Verena Heidrich-Meisner and Christian Igel. Neuroevolution Strategies for Episodic Reinforcement Learning. Journal of Algorithms 64(4), pp. 152-168, 2009

Multi-scale perception and action based on the free-energy principle

Stefan J. Kiebel (MPI für Kognitions- und Neurowissenschaften, DE)

We present a new theory about how an adaptive agent like the brain can perform prediction of sensory input and action online. We propose a Bayesian online inference scheme based on multi-scale temporal hierarchical dynamic systems. Research in robotics has shown that computing the selection of actions is difficult in a natural environment. We suggest that the conceptual framework and the computations can be simplified, under the free-energy principle. In this framework, action is a consequence of sensory prediction error at lowest level, i.e. the

fastest time-scale. As a consequence, the agent does not need to compute future sensory input caused by potentially appropriate actions, thereby enabling fast and robust responses. We will illustrate these points using simulations of multi-scale speech dynamics, oculomotor control, cued and goal-directed movements.

Keywords: Bayesian inference, free-energy principle, prediction, action, dynamic system

Learning to Control Dynamical Systems

Guenther Palm (Universität Ulm, DE)

After introducing basic control theory, I will describe the framework of adaptive critic design and our recent experiments with ESNs in this framework.

Keywords: Neural control, neurodynamic programming, echo-state networks

Task encoding and strategy learning in large worlds

Subramanian Ramamoorthy (University of Edinburgh, GB)

In recent years, we have seen increasing interest in the idea of robust autonomous agents that can operate reliably, over long periods of time, in dynamic environments involving both low-level sensorimotor uncertainties and large scale structural shifts. The question of how best to capture such dynamic phenomena in concise models that make decision making tractable is relatively open.

In this talk, I will outline an approach aimed at addressing some aspects of this issue. One way to accommodate the infinite variability of tasks and environments is to structure our models and strategies in terms of a hierarchy of abstractions, separating weak sufficient conditions from more variable quantitative information.

I will first explore this idea in the context of the problem of bipedal walking in humanoid robotics. I will show that there is interesting intrinsic structure in such tasks which can be exploited to craft layered representations enabling efficient data-driven learning. I will then show that even in environments with higher levels of nonstationarity, such as where costs/rewards, dynamics and goals are all continually changing, one can devise hierarchical strategies that exploit a similar separation of concerns. I will outline a game theoretic strategy learning algorithm for this case, along with a simple demonstration in an adversarial navigation example. I will then conclude with some remarks regarding how these ideas can be connected with recent developments in relational learning and other ways of exploiting structure in statistical learning.

Learning Machines

Martin Riedmiller (Universität Freiburg, DE)

Our research vision is to build intelligent machines that are finally able to acquire their capabilities completely from scratch. In this talk I will present some recent results of real world applications that both learn perception and decision making by reinforcement learning.

Keywords: Deep fitted q iteration, neural networks, reinforcement learning, neural fitted q iteration

Some Challenges of Robotics

Helge Ritter (Universität Bielefeld, DE)

Given the vivid evolution of available robot platforms, together with significant processing ressources, what are the achievements in the field and which challenges remain to be solved for further advances? State-of-the art examples of seemingly "simple" everyday actions, such as climbing, towel folding or bimanual multifingered jar opening reflect significant advances in the last decade, but also expose major gaps in our understanding and methods: pattern recognition methods need to be extended to interaction patterns, the current "understanding" of robots of their actions is extremely shallow and must be deepened, their representations of objects and interactions do not reach much beyond low-level aspects, we need to embrace more strongly "non-cartesian", non-metric and topological aspects.

The currently strongly physical "touch" of representations needs to be extended to account for agency, intentionality and even emotions. A fruitful theory field might be a taxonomoy and analysis of the different ways to create, shape and maintain structured relations at different levels of abstraction.

Such theory would have to be aided by good schemes for the integration of many heterogeneous functionalities in order to build artificial systems that are truly cognitive.

Keywords: Robotics, cognition, interaction pattern

Formal Theory of Fun and Creativity

Juergen Schmidhuber (IDSIA - Lugano, CH)

To build a creative agent that never stops generating non-trivial & novel & surprising data, we need two learning modules: (1) an adaptive predictor or compressor or model of the growing data history as the agent is interacting with its environment, and (2) a general reinforcement learner. The LEARNING PROGRESS of (1) is the FUN or intrinsic reward of (2).

That is, (2) is motivated to invent interesting things that (1) does not yet know but can easily learn. To maximize expected reward, in the absence of external reward (2) will create more and more complex behaviors that yield temporarily surprising (but eventually boring) patterns that make (1) quickly improve. We discuss how this principle explains science & art & music & humor, and how to scale up previous toy implementations of the theory since 1991, using recent powerful methods for (1) prediction and (2) reinforcement learning.

Keywords: Formal Theory of Fun and Creativity, science, art, music, humor

Full Paper:
http://www.idsia.ch/~juergen/interest.html

Analytical Functional Inductive Programming as Cognitive Rule Acquisition Device

Ute Schmid (Universität Bamberg, DE)

One of the most admirable characteristic of the human cognitive system is its ability to extract generalized rules covering regularities from example experience presented by or experienced from the environment. Humans' problem solving, reasoning and verbal behavior often shows a high degree of systematicity and productivity which can best be characterized by a competence level reflected by a set of recursive rules. While we assume that such rules are different for different domains, we believe that there exists a general mechanism to extract such rules from only positive examples from the environment. Our system Igor2 is an analytical approach to inductive programming which induces recursive rules by generalizing over regularities in a small set of positive input/output examples. We applied Igor2 to typical examples from cognitive domains and can show that the Igor2 mechanism is able to learn the rules which can best describe systematic and productive behavior in such domains.

Keywords: Inductive programming, rule learning, problem solving

Reservoir Computing

Benjamin Schrauwen (Gent University, BE)

Recurrent neural networks (RNNs) carry the promise of implementing efficient and biologically plausible signal processing, but they are notoriously hard to train. A compelling solution to the problem is to not train the internal recurrent connections of the network, but only a linear readout function. These techniques are generally termed Reservoir Computing and they are easy and efficient to train, and have a limited number of parameters to tune. They can furthermore be applied to a wide array of sequence processing tasks, and have already shown

state-of-the-art performance on several benchmark tasks. They constitute however more than an efficient way to train recurrent neural networks: the can serve as a "scaffold for learning", they embody a new paradigm to think about the implementation of computation, and they form a compelling "building block" for constructing multi-scale temporal hierarchies.

Keywords: Reservoir computing, recurrent neural networks

Goal Babbling for Efficient Explorative Learning of Inverse Kinematics

Jochen J. Steil (Universität Bielefeld, DE)

We present an approach to learn inverse kinematics of redundant systems from exploration without prior- or expert-knowledge.

The method allows for an iterative bootstrapping and refinement of the inverse kinematics estimate. The essential novelty lies in a path based sampling approach: we generate training data along paths, which result from execution of the currently learned estimate along a desired path towards a goal.

We derive and illustrate the exploration and learning process with a low-dimensional kinematic example that provides direct insight into the bootstrapping process. We further show that the method scales for high dimensional problems, such as the Honda humanoid robot or hyperredundant planar arms with up to 50 degrees of freedom. Finally we speculate that other inversion problems in cognitive domains face similar problems and could be efficiently tackled by similar exploratory methods.

One-shot Learning of Poisson Distributions in fast changing environments

Peter Tino (University of Birmingham, GB)

In Bioinformatics, Audic and Claverie were among the first to systematically study the influence of random fluctuations and sampling size on the reliability of digital expression profile data.

For a transcript representing a small fraction of the library and a large number N of clones, the probability of observing x tags of the same gene will be well-approximated by the Poisson distribution parametrised by its mean (and variance) $m>0$, where the unknown parameter m signifies the number of transcripts of the given type (tag) per N clones in the cDNA library.

On an abstract level, to determine whether a gene is differentially expressed or not, one has two numbers generated from two distinct Poisson distributions and based on this (extremely sparse) sample one has to decide whether the two Poisson distributions are identical or not. This can be used e.g. to determine equivalence of Poisson photon sources (up to time shift) in gravitational lensing.

Each Poisson distribution is represented by a single measurement only, which is, of course, from a purely statistical standpoint very problematic.

The key instrument of the Audic-Claverie approach is a distribution P over tag counts y in one library informed by the tag count x in the other library, under the null hypothesis that the tag counts are generated from the same but unknown Poisson distribution. P is obtained by Bayesian averaging (infinite mixture) of all possible Poisson distributions with mixing proportions equal to the posteriors (given x) under the flat prior over m.

We ask: Given that the tag count samples from SAGE libraries are *extremely* limited, how useful actually is the Audic-Claverie methodology? We rigorously analyse the A-C statistic P that forms a backbone of the methodology and represents our knowledge of the underlying tag generating process based on one observation.

We show will that the A-C statistic P and the underlying Poisson distribution of the tag counts share the same mode structure. Moreover, the K-L divergence from the true unknown Poisson distribution to the A-C statistic is minimised when the A-C statistic is conditioned on the mode of the Poisson distribution. Most importantly (and perhaps rather surprisingly), the expectation of this K-L divergence never exceeds 1/2 bit! This constitutes a rigorous quantitative argument, extending the previous empirical Monte Carlo studies, that supports the wide spread use of Audic-Claverie method, even though by their very nature, the SAGE libraries represent very sparse samples.

Keywords: Audic-Claverie statistic, Bayesian averaging, information theory, one-shot learning, Poisson distribution

Full Paper:
http://www.biomedcentral.com/1471-2105/10/310/abstract

See also: Basic properties and information theory of Audic-Claverie statistic for analyzing cDNA arrays, BMC Bioinformatics 2009, 10:310doi:10.1186/1471-2105-10-310

Basic properties and information theory of Audic-Claverie statistic for analyzing cDNA arrays

Peter Tino (University of Birmingham, GB)

The Audic-Claverie method [1] has been and still continues to be a popular approach for detection of differentially expressed genes in the SAGE framework. The method is based on the assumption that under the null hypothesis tag counts of the same gene in two libraries come from the same but unknown Poisson distribution. The problem is that each SAGE library represents only a single measurement. We ask: Given that the tag count samples from SAGE libraries are extremely limited, how useful actually is the Audic-Claverie methodology? We rigorously analyze the A-C statistic that forms a backbone of the methodology

and represents our knowledge of the underlying tag generating process based on one observation.

We show that the A-C statistic and the underlying Poisson distribution of the tag counts share the same mode structure. Moreover, the K-L divergence from the true unknown Poisson distribution to the A-C statistic is minimized when the A-C statistic is conditioned on the mode of the Poisson distribution. Most importantly, the expectation of this K-L divergence never exceeds 1/2 bit.

A rigorous underpinning of the Audic-Claverie methodology has been missing. Our results constitute a rigorous argument supporting the use of Audic-Claverie method even though the SAGE libraries represent very sparse samples.

Keywords: Audic-Claverie statistic, Bayesian averaging, information theory, one-shot learning, Poisson distribution

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2010/2799

Why deterministic logic is hard to learn but Statistical Relational Learning works

Marc Toussaint (TU Berlin, DE)

A brief note on why we think that the statistical relational learning framework is a great advancement over deterministic logic – in particular in the context of model-based Reinforcement Learning.

Keywords: Statistical relational learning, relational model-based Reinforcement Learning

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2010/2801

(Conceptual) Clustering for discovering Concept Drift and Concept Formation from Description Logics Knowledge Bases

Claudia d'Amato (University of Bari, IT)

Machine Learning is generally applied to knowledge bases that could be described by the use of several formalisms. In the real life, knowledge is generally changing over the time because new information is often available. Conceptual clustering methods can be used for learning new concepts from assertional knowledge, namely for learning novel concepts which are emerging based on the elicited clusters, and for detecting concept drift, that is concepts that are evolving, for instance because their intentional definitions do not entirely describe their extensions. The process can be structured in two steps. At the first step cluster

of elements are found. At the second step, supervised learning methods can exploit these clusters to induce new concept definitions or to refining existing ones. Moreover, intensionally defined groupings may be used for speeding-up searching tasks. In this talk, the focus will be on (conceptual) clustering methods for capturing concept drift and novelty detection for knowledge represented by means of ontologies which usually adopt (a fragment of) First Order Logic as a foundation of the representation language.

Different clustering approaches could be adopted. Here, the focus will be on a) an evolutionary approach that does not require the initial number of clusters; b) a fuzzy approach for allowing data belonging (with a certain degree of membership) to more than one cluster at the same time. In order to do this, effective similarity measures need to be used. For the purpose, a language-independent, semi-distance measure for individuals is adopted. It compares individuals with respect to a number of dimensions corresponding to a set of concept descriptions acting as a discriminating features set.

The measure is not absolute, it depends on the knowledge base it is applied to. Furthermore, in a setting where knowledge tends to evolve over the time, a reasonable assumption could be the *Open World Assumption* rather than the *Closed World Assumption* generally adopted in Machine Learning and Database contexts. This assumption needs to be taken into account for instance when similarity between individuals is computed.

Keywords: Clustering methods, concept drift and novelty detection, Open World Assumption Versus Closed World Assumption

Full Paper:
http://www.di.uniba.it/∼cdamato/eswc2008-PAM.pdf

See also:

References

1. F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, editors. *The Description Logic Handbook*. Cambridge University Press, 2003.
2. J.C. Bezdek and N.R. Pal. Some new indexes of cluster validity. *IEEE Transactions on Systems, Man, and Cybernetics*, 28(3):301–315, 1998.
3. A. Borgida. On the relative expressiveness of description logics and predicate logics. *Artificial Intelligence*, 82(1-2).
4. A. Borgida, T.J. Walsh, and H. Hirsh. Towards measuring similarity in description logics. In I. Horrocks, U. Sattler, and F. Wolter, editors, *Working Notes of the International Description Logics Workshop*, volume 147 of *CEUR Workshop Proceedings*, Edinburgh, UK, 2005.
5. C. d'Amato, N. Fanizzi, and F. Esposito. Reasoning by analogy in description logics through instance-based learning. In G. Tummarello, P. Bouquet, and O. Signore, editors, *Proceedings of Semantic Web Applications and Perspectives, 3rd Italian Semantic Web Workshop, SWAP2006*, volume 201 of *CEUR Workshop Proceedings*, Pisa, Italy, 2006.

6. C. d'Amato, S. Staab, N. Fanizzi, and F. Esposito. Efficient discovery of services specified in description logics languages. In T. Di Noia et al., editor, *Proceedings of Service Matchmaking and Resource Retrieval in the Semantic Web Workshop at ISWC 2007*, volume 243. CEUR, 2007.
7. M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases. In *Proceedings of the 2nd Conference of ACM SIGKDD*, pages 226–231, 1996.
8. N. Fanizzi, C. d'Amato, and F. Esposito. Induction of optimal semi-distances for individuals based on feature sets. In *Working Notes of the International Description Logics Workshop, DL2007*, volume 250 of *CEUR Workshop Proceedings*, Bressanone, Italy, 2007.
9. N. Fanizzi, C. d'Amato, and F. Esposito. Randomized metric induction and evolutionary conceptual clustering for semantic knowledge bases. In M. Silva, A. Laender, R. Baeza-Yates, D. McGuinness, O. Olsen, and B. Olstad, editors, *Proceedings of the ACM International Conference on Knowledge Management, CIKM2007*, Lisbon Portugal, 2007. ACM.
10. N. Fanizzi, L. Iannone, I. Palmisano, and G. Semeraro. Concept formation in expressive description logics. In J.-F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi, editors, *Proceedings of the 15th European Conference on Machine Learning, ECML2004*, volume 3201 of *LNAI*, pages 99–113. Springer, 2004.
11. A. Ghozeil and D.B. Fogel. Discovering patterns in spatial data using evolutionary programming. In John R. Koza, David E. Goldberg, David B. Fogel, and Rick L. Riolo, editors, *Genetic Programming 1996: Proceedings of the First Annual Conference*, pages 521–527, Stanford University, CA, USA, 1996. MIT Press.
12. L.O. Hall, I.B. Özyurt, and J.C. Bezdek. Clustering with a genetically optimized approach. *IEEE Trans. Evolutionary Computation*, 3(2):103–112, 1999.
13. S. Hirano and S. Tsumoto. An indiscernibility-based clustering method. In X. Hu, Q. Liu, A. Skowron, T. Y. Lin, R. Yager, and B. Zhang, editors, *2005 IEEE International Conference on Granular Computing*, pages 468–473. IEEE, 2005.
14. L. Iannone, I. Palmisano, and N. Fanizzi. An algorithm based on counterfactuals for concept learning in the semantic web. *Applied Intelligence*, 26(2):139–159, 2007.
15. A.K. Jain, M.N. Murty, and P.J. Flynn. Data clustering: A review. *ACM Computing Surveys*, 31(3):264–323, 1999.
16. L. Kaufman and P.J. Rousseeuw. *Finding Groups in Data: an Introduction to Cluster Analysis*. John Wiley & Sons, 1990.
17. J.-U. Kietz and K. Morik. A polynomial approach to the constructive induction of structural knowledge. *Machine Learning*, 14(2):193–218, 1994.
18. M. Kirsten and S. Wrobel. Relational distance-based clustering. In David Page, editor, *Proceedings of the 8th International Workshop, ILP98*, volume 1446 of *LNCS*, pages 261–270. Springer, 1998.
19. J. Lehmann. Concept learning in description logics. Master's thesis, Dresden University of Technology, 2006.
20. J. Lehmann and P. Hitzler. A refinement operator based learning algorithm for the alc description logic. In *the 17th International Conference on Inductive Logic Programming (ILP), 2007*, LNCS. Springer, 2007.
21. O. Nasraoui and R. Krishnapuram. One step evolutionary mining of context sensitive associations and web navigation patterns. In *Proceedings of the SIAM conference on Data Mining*, pages 531–547, Arlington, VA, 2002.
22. R. Ng and J. Han. Efficient and effective clustering method for spatial data mining. In *Proceedings of the 20th Conference on Very Large Databases, VLDB94*, pages 144–155, 1994.

23. S.-H. Nienhuys-Cheng. Distances and limits on herbrand interpretations. In D. Page, editor, *Proceedings of the 8th International Workshop on Inductive Logic Programming, ILP98*, volume 1446 of *LNAI*, pages 250–260. Springer, 1998.
24. Z. Pawlak. *Rough Sets: Theoretical Aspects of Reasoning About Data*. Kluwer Academic Publishers, 1991.
25. M. Sebag. Distance induction in first order logic. In S. Džeroski and N. Lavrač, editors, *Proceedings of the 7th International Workshop on Inductive Logic Programming, ILP97*, volume 1297 of *LNAI*, pages 264–272. Springer, 1997.
26. E.J. Spinosa, A. Ponce de Leon Ferreira de Carvalho, and J. Gama. OLINDDA: A cluster-based approach for detecting novelty and concept drift in data streams. In *Proceedings of the 22nd Annual ACM Symposium of Applied Computing, SAC2007*, volume 1, pages 448–452, Seoul, South Korea, 2007. ACM.
27. R. E. Stepp and R. S. Michalski. Conceptual clustering of structured objects: A goal-oriented approach. *Artificial Intelligence*, 28(1):43–69, Feb. 1986.
28. G. Widmer and M. Kubat. Learning in the presence of concept drift and hidden contexts. *Machine Learning*, 23(1):69–101, 1996.
29. P. Zezula, G. Amato, V. Dohnal, and M. Batko. *Similarity Search – The Metric Space Approach*. Advances in Database Systems. Springer, 2007.