Intelligenza Artificiale 7 (2013) 77–89
DOI 10.3233/IA-130050
IOS Press

Italian Machine Learning and Data Mining research: The last years

Nicola Di Mauro[a,*], Paolo Frasconi[b], Fabrizio Angiulli[c], Davide Bacciu[d], Marco de Gemmis[a], Floriana Esposito[a], Nicola Fanizzi[a], Stefano Ferilli[a], Marco Gori[e], Francesca A. Lisi[a], Pasquale Lops[a], Donato Malerba[a], Alessio Micheli[d], Marcello Pelillo[f], Francesco Ricci[g], Fabrizio Riguzzi[h], Lorenza Saitta[i] and Giovanni Semeraro[a]
[a]*University of Bari "Aldo Moro", Bari, Italy*
[b]*University of Florence, Florence, Italy*
[c]*University of Calabria, Rende, Italy*
[d]*University of Pisa, Pisa, Italy*
[e]*University of Siena, Siena, Italy*
[f]*University of Venice, Venice, Italy*
[g]*Free University of Bozen-Bolzano, Bolzano, Italy*
[h]*University of Ferrara, Ferrara, Italy*
[i]*University of Piemonte Orientale, Alessandria, Italy*

Abstract. With the increasing amount of information in electronic form the fields of Machine Learning and Data Mining continue to grow by providing new advances in theory, applications and systems. The aim of this paper is to consider some recent theoretical aspects and approaches to ML and DM with an emphasis on the Italian research.

Keywords: Machine Learning, Data Mining

1. Introduction

Theoretical aspects and approaches to Machine Learning (ML) and Data Mining (DM) continue to grow especially in these years where there is an increasing amount of data in electronic form. ML and DM are two closely related research fields that differ slightly in terms of their emphasis and terminology. In ML the emphasis is on developing methods able to automatically detect patterns in observed data and then to use them *to predict* future data or other outcomes of interest [1]. The objective of DM is to discover useful information and knowledge from a large collection of data. DM looks for interpretable models, whereas Machine Learning searches for accurate models. The aim of this paper is to report some recent theoretical aspects and approaches to ML and DM with an emphasis on the Italian research.

2. Multi-relational learning

Problems and research lines investigated at the Machine Learning group in Bari related to learning with complex representations ultimately based on logic languages are briefly presented in the following.

The growing diversification and complexity of the domains in which Machine Learning may find profitable application has recently brought much interest of the AI research community on techniques coming from the Inductive Logic Programming (ILP)

*Corresponding author: Nicola Di Mauro, University of Bari "Aldo Moro", Bari, Italy. E-mail: nicola.dimauro@uniba.it.

area [2]. However, the growing need of applying multi-relational techniques to real-world problems in which intelligent systems must interact with several kinds of users and provide continuous adaptation to their needs, goals and preferences has required to step up from the purely logical setting to a mixed one in which different approaches, strategies and representations must cooperate. Outstanding examples are the research efforts in Multi-Strategy Learning (MSL) and Statistical Relational Learning (SRL), that significantly extended the representational and inferential capabilities of ILP systems. SRL poses the problem of effective and efficient inference and structure learning algorithms that can be tackled by using metaheuristics [3]. In the same perspective, a particularly hot question is how to overcome the limitations of traditional *batch* approaches, in which all the relevant observations must be fully known at the time of learning, and any change in the set of observations involves withdrawing the knowledge acquired thus far and making hard efforts to restart from scratch. In fact, additional flexibility is needed to enable the systems to immediately react to new observations as long as they become available. An answer to this requirement is represented by the development of fully and inherently incremental systems, that can revise an existing theory whenever it proves unable to account for new evidence. Useful support may also come from the development of techniques for comparison and similarity assessment in Horn Clause logic, where the problem of indeterminacy in mapping (portions of) formulas adds further complexity. More advantages are ensured by bringing these two techniques to cooperation (e.g., as shown in [4]). Examples of successful applications include Document Image Understanding for Digital Library management and User profiling in Ambient Intelligence.

The perspective of the *Semantic Web* as a *Web of data* calls for learning methods that are able to deal with its shared semantics expressed with specific representations, that are ultimately based on Description Logics (DL). The main problems with dealing with such datasets arise from the scalability of reasoning w.r.t. its size and the inherent uncertainty of information in evolving Web-scale distributed scenarios (e.g. the *Linked Data* cloud). ML methods can provide effective solutions by means of an inductive approach. Since the 90s, methods have been proposed for learning DL-ontologies through the induction (refinement) of logical axioms for their terminological part in accordance to the available assertions. Generalization and covering relationships descend directly from the related proof-theory [5]. More recently, further methods have been proposed for learning statistical models enabling forms of approximate reasoning [6]. These models can be used to estimate the probability of statements on the instance-level, which are neither explicitly asserted in the knowledge base nor can be proven (or refuted) based on crisp logical reasoning. In particular, targeted methods range from those that exploiting semantic similarity measures and apply non-parametric learning, such as multilayer networks and kernel machines, up to statistical learning with relational representations, namely matrix/tensor decomposition, relational graphical models and first-order probabilistic approaches (see [7] for a recent survey). ML methods have been applied to enrich/extend ontologies on the schema level, supporting tasks like *ontology construction* (*ontology learning* from text), *evaluation*, *refinement*, *evolution*, as well as their *alignment*.

Another stream of research promoted by the Machine Learning group in Bari has been the extension of ILP from the original knowledge representation framework of Logic Programming to those hybrid formalisms, collectively referred to as *onto-relational rule languages*, that combine rules (notably, Datalog) and ontologies (represented with DLs). Following the seminal work of Rouveirol and Ventos [8], several ILP settings have been proposed for learning such onto-relational rules (see [9] for a survey). All these proposals aim at accommodating ontologies in ILP in a clear, elegant and well-founded manner by building upon established work in knowledge representation. These novel ILP settings have been successfully applied in Spatial Data Mining, Semantic Web Mining, and Ontology Evolution.

3. Probabilistic inductive logic programming

Probabilistic Inductive Logic Programming (PILP) is concerned with the induction of models that combine logic programming with probability theory [10] in order represent complex and uncertain relationship among the entities of the domain. Since the real world is often uncertain and structured, PILP has received an increased attention in the last decade.

The Italian community has been active in the field since its beginnings. The system Mr-SBC for learning naïve Bayes models from relational data has been proposed already in 2003 [11]. The authors then applied relational naïve Bayes to a spatial setting by inducing probabilistic classifiers from association rules [12].

A comparison of the approach of Mr-SBC with a more traditional ILP technique shows that the first is usually faster [13]. Mr-SBC was later upgraded to a semi-supervised system by iterating a k-NN based re-classification of labeled and unlabeled examples in order to identify borderline examples [14]

A long standing research direction in PILP is the combination of ILP with kernel methods of statistical learning. In [15] the authors proposed to use logic programs to generate traces corresponding to specific examples and kernels to quantify the similarity between traces. k-FOIL [16] combines the well-known ILP system FOIL with kernel methods. FOIL is used for searching relevant clauses to be used as features in standard kernel methods. This produces a dynamic propositionalization framework that is further investigated in [17] where experimental results show that kFOIL is able to handle domains of moderately large size and multi-class problems achieving advantages both in efficiency and accuracy. The language kLog was introduced for specifying logical and relational learning problems at a high level in a declarative way and was applied to natural language processing [18].

A different research direction in PILP involves learning models in a probabilistic logic language. Two main families of languages can be identified: those based on logic programming, such as ProbLog, and those based on first order logic such as Markov Logic Networks (MLNs). For the first type of languages, the system RIB [19] learns the parameters using the information bottleneck method, while EMBLEM [20, 21] uses a version of EM in which the expectations are computed directly using the data structures built for inference. Learning the structure and the parameters of unrestricted programs at the same time is performed by SLIPCASE [22] that searches the space of theories guided by the log-likelihood of the examples as the heuristic. For the second type of languages discriminative structure learning is achieved in [23] using local search methods. The use of metaheuristics for inference and learning is further investigated in [3].

Very recently PILP techniques have been applied to the Semantic Web: SRL methods are proposed in [7] for building ontologies from data, while a terminological naïve Bayesian classifiers is presented in [24].

4. Sub-symbolic learning in structured domains

The field of ML for Structured Domains (SD) deals with combinatorial data structures, such as sequences, trees and graphs. Structured data model aggregates of labeled elements and their intertwining relationships, providing fundamental representation and abstraction tools in computer science and AI, with applications e.g. in NLP, document analysis, network and Web data analysis, image processing, cheminformatics, neuroscience and bioinformatics. Supplying learning systems with the capability of processing SD in their full relational representation is a key to novel successful applications dealing with real-world complex data.

The topic has been approached by different ML paradigms, including SRL [25], relational DM (see e.g. [26] and [27] for a short recent overview), and kernel-based methods (see e.g. [27, 28] and the references therein). All these solutions share the common aim of overcoming the restrictions of traditional ML approaches, which ensue from assuming a flat vectorial encoding for the processed data. In this section, we briefly review the development results in the field from a sub-symbolic perspective, with a particular emphasis on the works of Italian research groups in the area.

Recursive Neural Networks (RNNs) have been, since their early onset, an elegant way to realize a learning system combining the flexibility and robustness of the connectionist (sub-symbolic) approach with the representational power of a structured domain. The basic idea is to follow through a state transition system the hierarchical/recursive nature of the data and hence to recursively compose the state encoding computed by a ML model for each vertex of a tree structure. Relevant contributions encompass both supervised approaches (classification/regression), see [29, 30], as well as unsupervised approaches that extend self organizing maps to SD, see [31] for a general framework and more references. Theoretical results on the model capabilities, e.g. showing the universal approximation property over tree domains, have been introduced in [32]. An extended survey and historical remarks on the introduction of RNNs can be found in [33]. The same recursive framework has been applied in the context of fuzzy systems [34], in a generative setting, resulting in hidden tree Markov models ([30] and more recently [35, 36] for efficient bottom-up approaches), as well as for developing efficient approaches within the reservoir computing field [37].

When moving to more complex structured data, contextual approaches have been showed to effectively overcome the constraint of recursive causal processing [38] and to extend the universal approximation capabilities to classes of directed positional acyclic graphs [39]. The extension to more general classes of

structured information, including both acyclic/cyclic, directed/undirected labeled graphs, has been pursued through different approaches, including the use of covering trees by RNNs [40], the construction of constrained dynamical transition systems [41], and the exploitation of contextual information in constructive neural models realizing a state transition systems without recursive dynamics [42].

The research on ML models for structured data has been paralleled by the development of a wide collection of impacting applications. A striking example is in the field of cheminformatics, where data structures are commonly used to represent molecular compounds. The generality of a SD learning has introduced benefits for the direct exploitation of the structural information conveyed by the molecular graph in the Quantitative Structure-Activity or Property Relationships analysis (see [43] for early approaches), with potential research impact on drug and biomaterial design, toxicology analysis, environmental and health studies.

Generally speaking, the ability to treat the inherent relational nature of the data in its fully-informative structured representation emerges as a key feature to further drive AI and ML methodologies. As such, learning in SD characterizes as an open research area with great interest from the perspective of the integration between symbolic and sub-symbolic learning, as well as for generalizing the ML theoretical basis to the SD. Nevertheless, the authors believe that such models already provide tools that can foster the development of effective and impacting real-world AI applications.

5. The complexity of relational learning

Relational Learning is a subfield of symbolic Machine Learning requiring the acquisition of knowledge from examples that are composed by interrelated parts, and hence cannot be represented by simple vectors of attribute values. Started by Michalski's pioneering work [44], the topic was developed very early in Italy [45, 46], and then, over the years, it evolved toward ILP [47], first, and SRL, thereafter [48].

Since the beginning, the main obstacle on the way to an effective use of relational learning has been its high computational complexity, primarily due to the *covering test*, i.e., the task of *matching* a candidate hypothesis φ (a logical formula with n variables) against an example x (a ground logical formula or a set of relational tables). This task has a complexity which is exponential in the number of variables in the hypothesis, and, during a learning session, it may have to be performed even thousands of times. No wonder, then, that many efforts have been invested in controlling this complexity. Two main ways have been followed toward this goal: either limiting the expressive power of the hypothesis language, or trying to reduce the learning problem to one in Propositional Logic.

According to the first approach, relational learners have often set strong biases on the hypothesis language [49, 50]. Well-known constraints are the *Object Identity* setting [51], the enforcement of *determinacy*, and the limitation of the variables' *depth*. The two latter can be combined to define *ij-determinacy*. Imposing determinacy limits both the complexity of the covering test, and the size of the hypothesis space.

Some formal results have been obtained within the PAC-learnability framework. For instance, Džeroski et al. [52] showed that non-recursive, constant-depth, determinate clauses are PAC-learnable. This result was extended by Cohen [53] to linear, closed, recursive, constant-depth determinate clauses. Also, *ij-clausal theories* were proved to be PAC-learnable [54]. An approach based on stochastic sampling has also been proposed [55]; it exhibits a polynomial complexity, but looses precision.

PAC-learnability, as well as classical complexity theory, is based on a worst-case analysis of a task. In trying to replace the notion of *worst case* complexity with that of *typical case* complexity, introduced in the analysis of SAT and CSP problems, Giordana and Saitta [56] have uncovered that the matching problem (the covering test) showed a *phase transition* with respect to the number m of predicates in the hypothesis φ and the number L of constants in the example x. A phase transition is a phenomenon that emerges in classes of randomly generated problems; in this case, a pair (φ, x) is built up according to a given stochastic model. The (m, L) plane is thus divided by a thin boundary (the phase transition) into two regions, one in which a randomly chosen φ almost surely "covers" x, and one in which it almost surely doesn't. Giordana et al. [56, 57] showed, experimentally, that when a learning problem is situated "near" the phase transition boundary, the computational requirement to perform a single matching becomes prohibitive even in very simple learning settings.

This discovery triggered interest in the phenomenon, especially in France and Germany, where researchers analyzed in the same light both relational and propositional learning [58–60]. Moreover, the framework has been extended to Grammar Induction [61], Multi-instance kernels in Support Vector Machine [62], and

to the very hypothesis search process [63, 64]. The most interesting outcome from these studies is that relational learning shows a double phase transitions, one in the covering test, and one in the hypothesis space exploration [63]. As a consequence, learning structural knowledge appears to be out of reach, if not for very simple cases.

The alternative approach, used to tame complexity, is *propositionalization*: the relational learning task is translated into one solvable by a propositional learner in polynomial time [65–67]. Stochastic approaches to propositionalization have also been suggested [68, 69]. Clearly, as the covering test is NP-hard, complete equivalence is impossible if P $\neq$ NP. Considering that also propositional learning may be affected by the emergence of a phase transition, certainly there are cases in which the transformation is ineffective.

More recently, with the diffusion of graphical models, relational learning underwent a deep renovation, entering as an essential part the field of SRL [48, 70]. Bayesian Networks, Markov Random Fields and Markov models are increasingly used to represent statistical relations among sets of variables. Relational learning, then, consists in finding both the structure and the parameters of a network. Unfortunately, also in this field both inference and learning are mostly intractable, and approximations are required [48]. As before, complexity is reduced either by limiting the possible structures of the network (reducing thus its expressive power) or by introducing various types of approximations and constraints [71–73]. Interestingly, some problematics of learning graphical models have been extended to mining social networks [74]; moreover, also Support Vector Machines and Neural Networks start to incorporate structured elements [75].

By comparing the initial approaches to relational learning with the set of computational tools available today, we cannot but wonder in front of the achieved results. Nevertheless, relation learning still remains a hard problem to be completely solved.

6. Temporal, spatial and spatio-temporal data mining

Temporal data are characterized by attributes whose values change with time, such as a geophysical property monitored by a sensor. Time implies an ordering which affects both the statistical properties of data and the semantics of the rules being extracted from them. Although the temporal dimension may provide better insights into data, it may also pose additional challenges when the statistical distribution associated to a property presents a time drift. The mining task can become even more difficult when data are generated in a continuous flow (or stream), eventually at high speeds, which prevents the collection of all the data in the memory before starting their analysis.

Spatial data are characterized by both a position and an extension in some space, which implicitly define many spatial relations, of various nature (directional, topological and distance). When *spatial autocorrelation* occurs, i.e. an attribute correlates with itself across space, there is a clear violation of one of the fundamental assumptions of classic data mining algorithms, that is, the independent generation of data samples. However, it is the identification of some form of spatial correlation which helps to clarify what the relevant spatial relationships are among the infinitely many that are implicitly defined by locational properties of spatial objects.

A further degree of complexity in temporal/spatial mining is introduced by the necessity of performing temporal/spatial reasoning to reach valid conclusions regarding the objects' relationships. Embedding temporal and spatial reasoning in data mining algorithms is crucial to make the right inferences either when patterns are generated or when patterns are evaluated.

The Italian research community for data mining and machine learning has been very active in the last years on the topics of temporal, spatial and spatio-temporal data mining. Far from being exhaustive, some of the most prominent results are reported below.

As to temporal data mining, most of the recent works have focused on the temporal dimension in networks. This interest is mainly due to the fact that in many application domains data naturally come in the form of a network, such as in social networks and sensor networks. The network is the unit of analysis which can evolve with time, and temporal data mining techniques are useful to reveal evolution-related patterns, such as eras and turning points [76], as well as evolution chains [77]. Relational approaches have also been investigated in order to properly deal with temporal relations between data received by a sensor network [78], as well as to analyze multidimensional time-series, known as longitudinal data [79]. A different perspective is offered by Loglisci and Ceci [80], who investigate how to discover a temporal relation, called bisociation, between concepts from two or more snapshots of a dynamic domain.

The relational approach has been advocated also for spatial data mining [81]. Indeed, relational mining

algorithms can be directly applied to various representations of networked data, i.e. collections of interconnected entities. By looking at spatial databases as a kind of networked data, where entities are spatial objects and connections are spatial relations, the application of relational mining techniques appears straightforward, at least in principle. Relational mining techniques can naturally take into account the various forms of correlation which bias learning in spatial domains. However, this capability is not exclusive of relational learning. For instance, predictive clustering trees (PCTs) have been extended in order to explicitly consider spatial autocorrelation in the values of the response (target) variable [82]. This extension seems particularly promising, since PCTs can be used for multi-target prediction.

The scarcity of labeled spatial data has driven researchers to investigate semi-supervised and transductive settings for spatial data mining. Indeed, these settings are based on a (semi-supervised) smoothness assumption, according to which if two points in a high-density region are close, then the corresponding predicted values at those points should also be close [83]. In spatial domains, where closeness of points corresponds to some spatial distance measure, this assumption is implied by (positive) spatial autocorrelation. Therefore, a strong spatial autocorrelation should counterbalance the lack of labeled data, when transductive relational learners [14] are applied to spatial domains. Results for spatial classification [84] and spatial regression tasks [85] support this expectation.

Most of the studies on mining spatio-temporal data generated by sensor networks consider only one dimension of the problem, usually the temporal dimension. However, this significantly limits the correct understanding of the problem. Ciampi et al. [86] have introduced a new kind of pattern, called trend cluster, to summarize a stream of spatial data. This summary is useful for many spatio-temporal data mining tasks, such as interpolation [87]. Following a similar approach, but related to moving objects, Monreale et al. [88] propose first extracting a concise representation, called trajectory pattern, of moving objects and then using it to forecast the next location of a moving object. The same type of patterns have been recently used to predict traffic congestions [89].

7. Learning from constraints

In the last few years, the research unit at the University of Siena has been involved in a new research field, referred to as *learning from constraints*. It focuses on the design of intelligent agents, centered around the parsimony principle, which are aimed to interact in complex environments, where sensorial data are combined with knowledge-based descriptions of the tasks. Unlike the classic framework of learning from examples, in those cases, the beauty and the elegance of simplicity behind the parsimony principle has not be profitably used for the formulation of systematic theories of learning yet. Most solutions are essentially based on hybrid systems, in which there is a mere combination of different modules that are separately charged of handling the prior knowledge on the tasks and of providing the inductive behavior naturally required in some tasks. The study of more unified approaches is not only of interest per se, but also, and perhaps primarily, because this crafting of knowledge with learning can give rise to interesting induction/deduction processes that are likely to be very effective in complex real-world problems.

In order to provide a unified context for manipulating perceptual data and granules of knowledge, we have proposed to use the unifying concept of *constraint.* It is sufficiently general to represent different kinds of sensorial data along with their relations, as well as to express abstract knowledge on the tasks. While the linguistic description to express a constraint can be of many different types, including those based on logic formalisms, in order to describe knowledge granules we can always end up into real-valued multi-variable functions involving the inputs and the learning tasks. We consider both the case in which we need perfect satisfaction (hard constraints) on a whole subset of the perceptual space and that in which only partially fulfillment is required (soft-constraints).

Examples of constraints come out naturally in different contexts: one might want to enforce the probabilistic normalization of a set of functions modeling a classification task, the probabilistic normalization of a density function, or might want to impose coherent decisions of the classifiers acting on different views of the same pattern. The expressive power of constraints becomes more significant when dealing with a specific problems in fields like vision, control, text classification, ranking in hyper-textual environment, and prediction of the stock market.

Interestingly, our notion of constraint, which is based on real-valued functions, encompasses logic predicates thanks to the classic connection established by the T-norm [90]. We unify continuous and discrete computational mechanisms so as to accommodate in the

same framework stimuli of very different kind, and propose the study of parsimonious agents interacting with constraints in a multi-task environment with the purpose of developing the simplest (smoothest) vectorial function in a set of feasible solutions. In a sense, our research naturally extend studies on approximation and learning presented in [91, 92] to the case in which the agent interacts with general hard/soft constraints instead of the classic interaction restricted to supervised examples. Interestingly, the case of supervised examples turns out to be a special case of the proposed framework and our theory reduces in that case to classic kernel machines. Moreover, while the optimization scheme restricted to a finite collection of supervised examples contains the traces of an induction process, when involving constraints with a significant degree of structure, their typical fulfillment does require to develop consistent solutions that are somehow related to deductive/abductive schemes. Recent research on learning from constraints can be found in [93–95] and, additional information on on sites.google.com/site/semanticbasedregularization/.

A remarkable consequence of the theory is the construction of a truly new inferential scheme, that we refer to as *parsimonious inference*, in which we bridge the gap between inferential schemes rooted in logic and in statistics. We carry out a parsimonious selection of the constraints to favor simple and elegant explanations. Preliminary studies can be found in [96].

8. Game-theoretic models in machine learning

The development of game theory in the early 1940's by von Neumann was a reaction against the view, dominant at that time, that problems in economic theory can be formulated using standard methods from optimization theory Indeed, most real-world economic problems typically involve conflicting interactions among decision-making agents that cannot be adequately captured by a single (global) objective function, thereby requiring a different, more sophisticated treatment. Accordingly, the main point made by game theorists is to shift the emphasis from optimality criteria to equilibrium conditions, namely to the search of a balance among multiple interacting forces. Interestingly, the later development of evolutionary game theory in the late 1970's by Maynard Smith [97] offered a dynamical systems perspective to game theory, an element which was totally missing in the traditional formulation, and provided powerful tools to deal with the equilibrium selection problem. As it provides an abstract theoretically-founded framework to elegantly model complex scenarios, game theory has found a variety of applications not only in economics and, more generally, in the social sciences but also in different fields of engineering and information technologies.

At the University of Venice, mainly within the EU FP7 SIMBAD project[1], game-theoretic concepts and tools to formulate several machine learning and pattern recognition problems, including data clustering [99, 100], structural matching [101], and semi-supervised learning [102] have been used. Indeed, these problems can naturally be formulated at an abstract level in terms of a game where (pure) strategies correspond to class labels and the payoff function is expressed in terms of competition between the hypotheses of class membership. These ideas and algorithms are finding applications in a variety of computer vision problems such as, e.g., object detection [103], motion analysis [104], and shape matching [105].

To illustrate the idea behind our approach, consider the (pairwise) clustering problem. Upon scrutinizing the relevant literature on the subject, it becomes apparent that the vast majority of the existing approaches deal with a very specific version of the problem, which asks for *partitioning* the input data into coherent classes. Instead of insisting on the idea of determining a partition of the input data, and hence obtaining the clusters as a by-product of the partitioning process, we reverse the terms of the problem and attempt instead to derive a rigorous formulation of the very notion of a cluster. This allows one, in principle, to deal with more general problems where clusters may overlap and/or clutter points may get unassigned.

The starting point of our approach is the elementary observation that a "cluster" may be informally defined as a maximally coherent set of data items, i.e., as a subset of the input data C which satisfies both an *internal* criterion (all elements belonging to C should be highly similar to each other) and an *external* one (all elements outside C should be highly dissimilar to the ones inside). We then define a "clustering game," and within this context we show that the notion of a cluster turns out to be equivalent to a classical equilibrium concept from (evolutionary) game theory (the evolutionary stable strategy [97]), as the latter reflects both the internal and external cluster conditions alluded to before. This characterization allows us to employ powerful dynamical systems from evolutionary game theory

[1] See http://simbad-fp7.eu and [98] for details.

such as the replicator dynamics (we refer to [100] for details).

One of the main attractive features offered by game theory, which distinguishes it from other approaches such as, e.g., spectral methods, is its generality, as it allows one to naturally deal with (dis)similarities that do not necessarily possess the Euclidean behavior or not even obey the requirements of a metric. Indeed, the lack of the Euclidean and/or metric properties undermines the very foundations of traditional pattern recognition theories and algorithms, and poses totally new theoretical/computational questions and challenges [98].

The classical approach to deal with non-geometric (dis)similarities is "embedding," which refers to any procedure that takes a set of (dis)similarities as input and produces a vectorial representation of the data as output, such that the proximities are either locally or globally preserved [106]. These approaches are all based on the assumption that the non-geometricity of similarity information can be eliminated or somehow approximated away. When this is not the case, i.e., when there is significant information content in the non-(geo)metricity of the data (see e.g., [107]), alternative approaches are needed, and game theory is particularly appealing in this respect as it makes no assumption whatsoever on the structure of the payoff (similarity) functions.

9. Outlier detection and description

Outlier detection and *description* concerns the quantitative and qualitative analysis of anomalies in data. Early methods for outlier identification have been developed in the field of *statistics* [108], but they assume that the given dataset has a distribution model. In order to overcome this limitation, in the last twenty years *data mining* and *machine learning* researchers have focused on *semi-supervised* (also called *one-class classification* or *novelty detection*) and *unsupervised* techniques [109]. In the former case only examples of the normal class are available, while in the latter the dataset is unlabelled and the objects most deviating from the remainder of the data are to be detected.

Most of the effort in these fields of the Italian community have been made by the data and knowledge engineering group of the DIMES Department of University of Calabria (former DEIS). In the following these contributions are overviewed in the context of the related literature.

In the *supervised* setting, *distance-based* [110] and *density-based* [111] approaches have gained popularity over the years and are today recognized among the most effective techniques for identifying outliers. Distance-based outliers are defined in terms of the number of objects lying in a fixed-radius neighborhood of each object. While this definition does not make any assumption on the data distribution, the first proposed algorithms ran in time quadratic in the size or exponential in the dimensionality of the data [110]. Thus, efforts for developing scalable algorithms have been subsequently made. In this scenario, contributions concern the design of *efficient algorithms* [112, 113].

Specifically [112] introduced a novel definition of distance-based outlier together with an algorithm called *HilOut*. The novel definition ranks objects on the basis of the sum of the distances separating each object from its k nearest neighbors and has become one of the definitions commonly adopted in applications. *HilOut* makes use of the Hilbert space-filling curve to deal with high-dimensional data, by mapping the d-dimensional Euclidean space onto the real segment $[0, 1]$. As a main merit, this is the first method guaranteeing an approximate solution within linear time in the dataset size and also scalable on high-dimensional data. The DOLPHIN algorithm [113] has been designed to work with disk resident datasets and detects outliers according to the definition in [110]. The strong points of this techniques are the very low I/O cost, the small theoretical memory usage, and the linear time performance with respect to the dataset size. These characteristics have established DOLPHIN has a state of the art outlier detection algorithm.

Distance and density based definitions are mainly considered in the context of unsupervised outlier detection, but can also be exploited as *semi-supervised* methods. Noteworthy contributions in this context concern the definition of *compressed representations* of the data for prediction purposes [114–116]: [114] introduced the concept of *outlier detection solving set* for the definition in [112], while [115, 116] generalized the above approach in order to encompass the definition provided in [110] to further reduce the size of the compressed set.

While traditional outlier analysis techniques have focused on the outlier detection task, recently the *outlier description* task has raised attention due to the need of understanding the reasons that make an object exceptional. This has connections with *subspace outlier mining*, where the attention is restricted to a subset of the features [117, 118]. Within this setting, the concept of outlier property and outlier explanation have been introduced in the context of cat-

egorical data in [119]: "Assume you are provided with the information that one of the individuals in a data population is abnormal, but no reason whatsoever is given to you as to why this particular individual is to be considered abnormal. The goal is to discover sets of attributes that account for the abnormality of that individual". In particular, a set of attributes is exceptional (i.e., an *outlier property*) if the occurrence frequency of the combination of values assumed by the outlier on these attributes is infrequent. An *outlier explanation* is a set of attributes that allows to select a subset of the data population containing the outlier and making another set of attributes exceptional. In [120] the above approach is extended to the case in which a *group of anomalous individuals* is given in input. This method resorts to a form of minimum distance estimation for evaluating the badness of fit of the values assumed by the outliers compared to the probability distribution associated with the values assumed by the inliers.

10. Recommender systems

Recommender Systems (RSs) are information search and filtering tools that provide suggestions for items to be of use to a user [121]. They are common in a large number of Internet applications, helping users to make better choices while searching for news, music, vacations or movies. "Item" is the general term used to denote what the system recommends to its users, and a specific RS normally focuses on one type of items.

A RS implements a real valued function of the product space of the users and items $r^* : U \times I \rightarrow \mathcal{R}$; $r^*(u, i)$ is the prediction of how a user $u \in U$ evaluates an item $i \in I$. Having a collection of predicted evaluations, a RS recommends to u the items i with the largest predicted evaluations $r^*(u, i)$ (top-N recommendations). Item evaluations are called ratings in Collaborative Filtering RSs; this is a popular technique that leverages the ratings of users estimated to be similar to the target user [121]. In some RSs more complex types of evaluations are observed, for instance the relative preference for an item when compared to another one [122], or the time spent by a user interacting with items [123].

Content-based RSs (CBRSs) [124], differently from Collaborative Filtering RSs, strongly rely on content descriptions of items in order to identify items similar to those the target user liked and therefore predicted to have a large user evaluation $r^*(u, i)$.

Novel research works have introduced semantic indexing techniques that shift from a keyword-based to a concept-based representation of items. Some semantic techniques allow to generate better recommendations by modelling user preferences with concepts defined in external knowledge bases [125]. To this end, the infusion of common-sense knowledge available in Open knowledge sources, such as Wikipedia, DBpedia and Freebase, have been shown to be useful to improve the indexing effectiveness of both item descriptions and user profiles [126].

Other semantic approaches, called distributional, build high dimensional semantic spaces in which words and concepts with similar meanings are close to each another (geometric metaphor of meaning) [127]. One of the great virtues of distributional approaches is that semantic spaces can be built using entirely unsupervised analysis of free text. In addition, replacing words with item descriptions results in a high dimensional space where similar items are represented close to one another [128].

Context-aware RSs (CARS) is a novel and promising subject. Here the RS tailors the recommendations to the specific contextual situations that may be experienced by the user while interacting with an item [129]. For instance [130] describes a place of interests (POIs) RS where fourteen contextual factors are considered, such as, the time of the travel, the weather, the user mood or the group that is accompanying the traveller.

Recent CARS techniques have addressed: the design of evaluation/rating prediction models that effectively incorporate the additional information brought by the context [131, 132]; methodologies for identifying and acquiring relevant contextual knowledge [130]; and applications dealing with particular contextual data, e.g., the current location of the user to adapt the music to the surrounding while the user is visiting a place or driving a car [133, 134]

A particular type of context is the group of users that experience an item together and may have conflicting preferences. Recent results have shown that rank aggregation techniques, previously developed for building meta search engines, can be effectively applied to group recommendation [135].

Evaluation of recommender systems has been performed with offline and online methods. Offline evaluations are based on standard cross validation techniques and error metrics (Root Mean Square Error). However [136] have shown that error metrics are not suited for assessing the system performance on top-N

recommendations, and accuracy metric, such as precision and recall, should be used.

Ultimately, offline evaluations can never assess the true performance of a RS and on line evaluations are in order. Cremonesi et al. focuses on the persuasive role of RSs and shows that algorithmic attributes are less crucial than expected in determining the user's perception of a RS quality, hence HCI methods must be considered [137]. Moreover, offline evaluations do not factor in the cost of user preferences elicitation, which has in practice an important role in determining the amount and quality of the data that a RS can leverage [138]. Hence, the usage of active learning techniques for user preference elicitation in RSs have been proposed [139].

Contributions

Sec. 2 was written by F. Esposito, N. Fanizzi, S. Ferilli and F.A. Lisi. Sec. 3 was written by F. Riguzzi. Sec. 4 was written by D. Bacciu and A. Micheli. Sec. 5 was written by L. Saitta. Sec. 6 was written by D. Malerba. Sec. 7 was written by M. Gori. Sec. 8 was written by M. Pelillo. Sec. 9 was written by F. Angiulli. Sec. 10 was written by M. de Gemmis, P. Lops, F. Ricci and G. Semeraro.

References

[1] K. Murphy, *Machine learning: A probabilistic perspective*, MIT Press, 2012.

[2] S. Muggleton, L. De Raedt, et al., Ilp turns 20 – biography and future challenges, *Machine Learning* **86**(1) (2012), 3–23.

[3] M. Biba, S. Ferilli, et al., Boosting learning and inference in Markov logic through metaheuristics, *Appllied Intelligence* **34**(2) (2011), 279–298.

[4] S. Ferilli, T.M.A. Basile, et al., A general similarity framework for horn clause logic, *Fundamenta Informaticae* **90**(1-2) (2009), 43–66.

[5] J. Lehmann, P. Hitzler, Concept learning in description logics using refinement operators, *Machine Learning* **78**(1-2) (2010), 203–250.

[6] F. Esposito, C. d'Amato, et al., Fuzzy clustering for semantic knowledge bases, *Fundam Informat* **99**(2) (2010), 187–205.

[7] A. Rettinger, U. Lösch, et al., Mining the semantic web – statistical learning for next generation knowledge bases, *Data Mining and Knowledge Discovery* **24**(3) (2012), 613–662.

[8] C. Rouveirol, V. Ventos, Towards learning in carin-aln. In J. Cussens and A. Frisch (editors), *Inductive Logic Programming*, volume 1866 of *LNCS*, Springer, 2000, pp. 191–208.

[9] F.A. Lisi, F. Esposito, On ontologies as prior conceptual knowledge in inductive logic programming, In *Knowledge Discovery Enhanced with Semantic and Social Information*, Springer, 2009, pp. 3–17.

[10] L. De Raedt, P. Frasconi, et al., (editors). *Probabilistic Inductive Logic Programming – Theory and Applications*, Springer, 2008.

[11] M. Ceci, A. Appice, et al., Mr-SBC: A multi-relational naïve Bayes classifier, In *ECML/PKDD '03*, volume 2838 of *LNCS*, Springer, 2003, pp. 95–106.

[12] M. Ceci, A. Appice, et al., Spatial associative classification at different levels of granularity: A probabilistic approach, In *ECML/PKDD '04* (2004), 99–111.

[13] M. Ceci, M. Berardi, et al., Relational learning: Statistical approach versus logical approach in document image understanding, In *AI*IA '05*, Springer, 2005, pp. 418–429.

[14] D. Malerba, M. Ceci, et al., A relational approach to probabilistic classification in a transductive setting, *Engineering Applications of Artificial Intelligence* **22**(1) (2009) 109–116.

[15] A. Passerini, P. Frasconi, et al., Kernels on Prolog proof trees: Statistical learning in the ILP setting, *J. Mach. Learn. Res.* **7** (2006), 307–342.

[16] N. Landwehr, A. Passerini, et al., kFOIL: Learning simple relational kernels. In *AAAI '06*, AAAI Press, 2006, pp. 389–394.

[17] N. Landwehr, A. Passerini, et al., Fast learning of relational kernels, *Machine Learning* **78**(3) (2010), 305–342.

[18] M. Verbeke, P. Frasconi, et al., Kernel-based logical and relational learning with kLog for hedge cue detection, In *ILP '12*, volume 7207 of *LNCS*, Springer, 2012, pp. 347–357.

[19] F. Riguzzi, N. Di Mauro, Applying the information bottleneck to statistical relational learning, *Machine Learning* **86**(1) (2012), 89–114.

[20] E. Bellodi, F. Riguzzi, Experimentation of an expectation maximization algorithm for probabilistic logic programs, *Intelligenza Artificiale* **8**(1) (2012), 3–18.

[21] E. Bellodi, F. Riguzzi, Expectation maximization over binary decision diagrams for probabilistic logic programs, *Intelligent Data Analysis* **17**(2) (2013), 343–363.

[22] E. Bellodi, F. Riguzzi, Learning the structure of probabilistic logic programs, In *ILP '12*, Springer, 2012, pp. 61–75.

[23] M. Biba, S. Ferilli, et al., Discriminative structure learning of Markov logic networks, In *ILP '08* (2008), 59–76.

[24] P. Minervini, C. d'Amato, et al., Learning probabilistic description logic concepts: Under different assumptions on missing knowledge, In *SAC '12*, ACM, 2012, pp. 378–383.

[25] L. Getoor, B. Taskar, *Introduction to Statistical Relational Learning*, The MIT Press, 2007.

[26] S. Dzeroski, N. Lavrac, *Relational Data Mining*, Springer-Verlag, 2001.

[27] G. Da San Martino, A. Sperduti, Mining structured data, *IEEE Comp. Int. Mag.* **5**(1) (2010), 42–49.

[28] B. Hammer, C. Saunders, et al., Special issue on neural networks and kernel methods for structured domains, *Neural Netw.* **18**(8) (2005), 1015–1018.

[29] A. Sperduti, A. Starita, Supervised neural networks for the classification of structures, *IEEE Trans. Neur. Netw.* **8**(3) (1997), 714–735.

[30] P. Frasconi, M. Gori, et al., A general framework for adaptive processing of data structures, *IEEE Trans. Neur. Netw.* **9**(5) (1998), 768–786.

[31] B. Hammer and A. Micheli, et al., Recursive self-organizing network models, *Neural Netw.* **17**(8-9) (2004), 1061–1085.

[32] B. Hammer, *Learning with Recurrent Neural Networks*, volume 254 of *LNCIS*, Springer, 2000.

[33] P. Frasconi, M. Gori, et al., From sequences to data structures: Theory and applications, In *A Field Guide to Dynamical Recurrent Networks*, Wiley-IEEE Press, 2001, pp. 351–374.

[34] M. Gori, A. Petrosino, Encoding nondeterministic fuzzy tree automata into recursive neural networks, *IEEE Trans. Neur. Netw.* **15**(6) (2004), 1435–1449. ISSN 1045-9227.

[35] D. Bacciu, A. Micheli, et al., Compositional generative mapping for tree-structured data – part i: Bottom-up probabilistic modeling of trees, *IEEE Trans. Neur. Netw. Learn. Sys.* **23**(12) (2012), 1987–2002.

[36] D. Bacciu, A. Micheli, et al., Compositional generative mapping for tree-structured data – part ii: Topographic projection model, *IEEE Trans. Neur. Netw. Learn. Sys.* **24**(2) (2013), 231–247.

[37] C. Gallicchio, A. Micheli, Tree echo state networks, *Neurocomp.* **101** (2013), 319–337.

[38] A. Micheli, D. Sona, et al., Contextual processing of structured data by recursive cascade correlation, *IEEE Trans. Neur. Netw.* **15**(6) (2004), 1396–1410.

[39] B. Hammer, A. Micheli, et al., Universal approximation capability of cascade correlation for structures, *Neural Comp.* **17**(5) (2005), 1109–1159.

[40] M. Bianchini, M. Gori, et al., Recursive processing of cyclic graphs, *IEEE Trans. Neur. Netw.* **17**(1) (2006), 10–18.

[41] F. Scarselli, M. Gori, et al., The graph neural network model, *IEEE Trans. Neur. Netw.* **20**(1) (2009), 61–80.

[42] A. Micheli, Neural network for graphs: A contextual constructive approach, *IEEE Trans. Neur. Netw.* **20**(3) (2009), 498–511.

[43] A.M. Bianucci, A. Micheli, et al., Application of cascade correlation networks for structures to chemistry, *Appl. Intell.* **12**(1-2) (2000), 115–145.

[44] R. Michalski, A theory and methodology of inductive learning, *Artificial Intelligence* **20** (1983), 83–134.

[45] F. Bergadano, A. Giordana, et al., Automated concept acquisition in noisy environments, *IEEE PAMI* **10** (1988), 555–578.

[46] F. Esposito, D. Malerba, et al., Flexible matching for noisy structural descriptions, In *IJCAI '91* (1991), 658–664.

[47] S. Muggleton (editor), *Inductive Logic Programming*, Academic Press, London, UK, 1992.

[48] D. Koller, N. Friedman, *Probabilistic Graphical Models*, The MIT Press, London, UK, 2009.

[49] J. Kietz, K. Morik, A polynomial approach to the constructive induction of structural knowledge, *Machine Learning* **14** (1994), 193–218.

[50] M. desJardins, D. Gordon (editors), *Machine Learning: Special Issue on Bias Evaluation and Selection*, volume 20, Kluwer Academic, 1995.

[51] F. Esposito, N. Fanizzi, et al. Theory refinement under Object Identity, In *ICML '00* (2000), 263–270.

[52] S. Džeroski, S. Muggleton, et al., PAC-learnability of determinate logic programs. In *COLT '92* (1992), 128–134.

[53] W. Cohen, Pac-learning recursive logic programs: Negative result, *J. Artif. Intel. Res.* **2** (1995), 541–573.

[54] L. De Raedt, S. Džeroski, First-order jk-clausal theories are PAC-learnable, *Artificial Intelligence* **70** (1994), 375–392.

[55] M. Sebag, C. Rouveirol, Tractable induction and classification in first order logic via Stochastic Matching, In *IJCAI '97* (1997), 888–893.

[56] A. Giordana, L. Saitta, Phase transitions in relational learning, *Machine Learning* **41**(2) (2000), 17–251.

[57] M. Botta, A. Giordana, et al., Relational learning as search in a critical region, *J. Mach. Learn. Res.* **4** (2003), 431–463.

[58] U. Rückert, S. Kramer, et al., Phase transitions and stochastic local search in k-term dnf learning, *LNCS* **2430** (2002), 43–63.

[59] N. Baskiotis, M. Sebag, C4.5 competence map: A phase transition-inspired approach, In *ICML '04* (2004), 73–80.

[60] J. Demongeot, C. Jézéquel, et al., Boundary conditions and phase transitions in neural networks, Theoretical results. *Neural Networks* **21** (2008), 971–979.

[61] N. Pernot, A. Cornuéjols, et al., Phase transitions within grammatical inference, In *Proc. 19th Int. Joint Conf. on Artificial Intelligence*, Edinburgh, Scotland, UK, 2005, pp. 811–816.

[62] R. Gaudel, M. Sebag, et al., A phase transition-based perspective on multiple instance kernels, *LNCS* **4894** (2008), 112–121.

[63] E. Alphonse, A. Osmani, On the connection between the phase transition of the covering test and the learning success rate in ilp, *Machine Learning* **70** (2008), 135–150.

[64] E. Alphonse, A. Osmani, Empirical study of relational learning algorithms in the phase transition framework, *LNCS* **5781** (2009), 51–66.

[65] S. Kramer, N. Lavrac, et al., Propositionalization approaches to relational data mining, In *Relational Data Mining*, Springer Verlag, 1992, pp. 262–291.

[66] J. Zucker, J. Ganascia, Representation changes for efficient learning in structurel domains, In *ICML '96* (1996), 543–551.

[67] E. Alphonse, C. Rouveirol, Lazy propositionalisation for relational learning, In *Proc. ECAI 2000* (2000), 256–260.

[68] S. Kramer, B. Pfahringer, et al., Stochastic propositionalization of non-determinate background knowledge, In *Proc. of the 8th Int. Conf. on ILP*, Prague, Czech Republic, 1997, pp. 80–94.

[69] N. Di Mauro, T.M.A. Basile, et al., Stochastic propositionalization for efficient multi-relational learning, In *Proc. 17th Int. Conf. on Found. of Intel. Sys.* (2008), 78–83.

[70] C. Bishop (editor), *Pattern Recognition and Machine Learning*, Springer, New York, NY, 2002.

[71] R. Gens, P. Domingos, Discriminative learning of sumproduct networks, In *Proc. NIPS*, (2012), 3248–3256.

[72] N. Lao, J. Zhu, et al., Efficient relational learning with hidden variable detection, In *Proc. NIPS'10* (2010), 1234–1242.

[73] C. Taranto, N.D. Mauro, et al., Learning in probabilistic graphs exploiting language-constrained patterns, *LNAI* **7765** (2013), 155–169.

[74] F. Esposito, S. Ferilli, et al., Social networks and statistical relational learning: A survey, *Int. J. of Social Network Mining* **1** (2012), 185–208.

[75] D. Bacciu, A. Micheli, et al., A generative multiset kernel for structured data, In *Proc. ICANN* (2012), 57–64.

[76] M. Berlingerio, M. Coscia, et al., Evolving networks: Eras and turning points, *Intell. Data Anal.* **17**(1) (2013), 27–48.

[77] C. Loglisci, M. Ceci, et al., Discovering evolution chains in dynamic networks, In *NFMCP*, volume 7765 of *LNCS*, Springer, 2012, pp. 185–199.

[78] T.M.A. Basile, N.D. Mauro, et al., Relational temporal data mining for wireless sensor networks, In *AI*IA*, volume 5883 of *LNCS*, Springer, 2009, pp. 416–425.

[79] C. Loglisci, M. Ceci, et al., A temporal data mining framework for analyzing longitudinal data, In *DEXA (2)*, volume 6861 of *LNCS*, Springer, 2011, pp. 97–106.

[80] C. Loglisci, M. Ceci, Discovering temporal bisociations for linking concepts over time, In *ECML/PKDD (2)*, volume 6912 of *LNCS*, Springer, 2011, pp. 358–373.

[81] D. Malerba, A relational perspective on spatial data mining, *IJDMMM* **1**(1) (2008), 103–118.

[82] D. Stojanova, M. Ceci, et al., Dealing with spatial autocorrelation when learning predictive clustering trees, *Ecological Informatics* **13** (2013), 22–39.

[83] O. Chapelle, B.S.B. et al. (editors), *Semi-supervised learning*, MIT Press, Cambridge: MA, 2006.

[84] M. Ceci, A. Appice, et al., Transductive learning for spatial data classification, In *Advances in Machine Learning I*, volume 262 of *SCI*, Springer, 2010, pp. 189–207.

[85] A. Appice, M. Ceci, et al., Transductive learning for spatial regression with co-training, In *SAC*, ACM, 2010, pp. 1065–1070.

[86] A. Ciampi, A. Appice, et al., Summarization for geographically distributed data streams, In *KES (3)*, volume 6278 of *LNCS*, Springer, 2010, pp. 339–348.

[87] A. Appice, A. Ciampi, et al., Using trend clusters for spatiotemporal interpolation of missing data in a sensor network, *J. Spatial Information Science* **6**(1) (2013), 119–153.

[88] A. Monreale, F. Pinelli, et al., Wherenext: A location predictor on trajectory pattern mining, In *KDD*, ACM, 2009, pp. 637–646.

[89] F. Giannotti, M. Nanni, et al., Unveiling the complexity of human mobility by querying and mining massive trajectory data, *VLDB J.* **20**(5) (2011), 695–719.

[90] E.P. Klement, R. Mesiar, et al., *Triangular norms*, Kluwer Academic Publishers Dordrecht, 2000.

[91] T. Poggio, F. Girosi, A theory of networks for approximation and learning, Technical report, DTIC Document, 1989.

[92] G. Gnecco, M. Gori, et al., Learning with boundary conditions, *Neural computation* **25**(4) (2013), 1029–1106.

[93] S. Melacci, M. Gori, Unsupervised learning by minimal entropy encoding, *IEEE Trans. Neural Netw. Learning Syst.* **23**(12) (2012), 1849–1861.

[94] M. Diligenti, M. Gori, et al., Bridging logic and kernel machines, *Machine learning* **86**(1) (2012), 57–88.

[95] S. Melacci, M. Gori, Learning with box kernels, *IEEE Trans. on Patt. Anal. and Mach. Intell.* **99** (2013), 1.

[96] M. Gori, S. Melacci, Constraint verification with kernel machines, *IEEE Trans. Neural Netw. Learning Syst.* **24**(5) (2013), 825–831.

[97] J.M. Smith, *Evolution and the Theory of Games*, Springer, 1993.

[98] M. Pelillo (editor), *Similarity-Based Pattern Analysis and Recognition*, Springer, 2013.

[99] M. Pavan, M. Pelillo, Dominant sets and pairwise clustering, *IEEE Trans. PAMI* **29**(1) (2007), 167–172.

[100] S.R. Bulò, M. Pelillo, A game-theoretic approach to hypergraph clustering, *IEEE Trans. PAMI* **35**(6) (2013), 1312–1327.

[101] A. Albarelli, S.R. Bulò, et al., Matching as a non-cooperative game, In *ICCV '09* (2009), 1319–1326.

[102] A. Erdem, M. Pelillo, Graph transduction as a noncooperative game, *Neural Computation* **24**(3) (2012), 700–723.

[103] P. Kontschieder, S.R. Bulò, et al., Evolutionary hough games for coherent object detection, *Computer Vision and Image Understanding*, 2012.

[104] A. Albarelli, E. Rodolà, et al., Imposing semi-local geometric constraints for accurate correspondences selection in structure from motion: A game-theoretic perspective, *International journal of computer vision* **97**(1) (2012), 36–53.

[105] E. Rodola, A.M. Bronstein, et al., A game-theoretic approach to deformable shape matching, In *CVPR '12* (2012), 182–189.

[106] E. Pękalska, R.P. Duin, *The dissimilarity representation for pattern recognition: foundations and applications*, 64, World Scientific, 2005.

[107] D.W. Jacobs, D. Weinshall, et al., Classification with nonmetric distances: Image retrieval and class representation, *IEEE Trans. PAMI* **22**(6) (2000), 583–600.

[108] V. Barnett, T. Lewis, *Outliers in statistical data*, John Wiley & Sons, 1994.

[109] V. Chandola, A. Banerjee, et al., Anomaly detection: A survey, *ACM Comput. Surv.* **41**(3) (2009), 1–58.

[110] E.M. Knorr, R.T. Ng, A unified notion of outliers: Properties and computation, In *KDD* (1997), 219–222.

[111] M.M. Breunig, H.-P. Kriegel, et al., Lof: Identifying density based local outliers, In *SIGMOD Conference* (2000), 93–104.

[112] F. Angiulli, C. Pizzuti, Outlier mining in large high-dimensional data sets, *IEEE Trans. Knowl. Data Eng.* **17**(2) (2005), 203–215.

[113] F. Angiulli, F. Fassetti, Dolphin: An efficient algorithm for mining distance-based outliers in very large datasets, *TKDD* **3**(1) (2009), 1–57.

[114] F. Angiulli, S. Basta, et al., Distance-based detection and prediction of outliers, *IEEE Trans. Knowl. Data Eng.* **18**(2) (2006), 145–160.

[115] F. Angiulli, Condensed nearest neighbor data domain description, *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(10) (2007), 1746–1758.

[116] F. Angiulli, Prototype-based domain description for one-class classification, *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(6) (2012), 1131–1144.

[117] C.C. Aggarwal, P.S. Yu, Outlier detection for high dimensional data, In *SIGMOD Conference* (2001), 37–46.

[118] J. Zhang, H.H. Wang, Detecting outlying subspaces for high-dimensional data: The new task, algorithms, and performance, *Knowl. Inf. Syst.* **10**(3) (2006), 333–355.

[119] F. Angiulli, F. Fassetti, et al., Detecting outlying properties of exceptional objects, *ACM T. Database Syst.* **34**(1) (2009), 1–62.

[120] F. Angiulli, F. Fassetti, et al., Discovering characterizations of the behavior of anomalous subpopulations, *IEEE Trans. Knowl. Data Eng.*, **25**(6) (2013), 1280–1292.

[121] F. Ricci, L. Rokach, et al., *Recommender Systems Handbook*, Springer, 2011.

[122] H. Blanco, F. Ricci, Inferring user utility for query revision recommendation, In *Proceedings of the 28th Symposium On Applied Computing ACM*, **1** (2013), 245–252.

[123] O. Moling, L. Baltrunas, et al. Optimal radio channel recommendations with explicit and implicit feedback, In *RecSys '12* (2012), 75–82.

[124] P. Lops, M. de Gemmis, et al., Content-based recommender systems: State of the art and trends, In *Recommender Systems Handbook*, Springer, 2011, pp. 73–105.

[125] G. Semeraro, M. Degemmis, et al., Combining learning and word sense disambiguation for intelligent user profiling, In *IJCAI '07* (2007), 2856–2861.

[126] G. Semeraro, P. Lops, et al., Knowledge infusion into content based recommender systems, In *RecSys '09* (2009), 301–304.

[127] M. Sahlgren, *The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*, Ph.D. thesis, Stockholm University, Department of Linguistics, 2006.

[128] C. Musto, F. Narducci, et al., Cross-language information filtering: Word sense disambiguation vs. distributional models, In *AI*IA 2011*, volume 6934 of *LNCS* (2011), 250–261.

[129] G. Adomavicius, B. Mobasher, et al., Context-aware recommender systems, *AI Magazine* **32**(3) (2011), 67–80.

[130] L. Baltrunas, B. Ludwig, et al., Context relevance assessment and exploitation in mobile recommender systems, *Personal and Ubiquitous Computing* **16**(5) (2012), 507–526.

[131] L. Baltrunas, B. Ludwig, et al., Matrix factorization techniques for context aware recommendation, In *RecSys '11*, ACM, 2011, pp. 301–304.

[132] L. Baltrunas, F. Ricci, Context-based splitting of item ratings in collaborative filtering, In *RecSys '09*, ACM Press, October 22–25, New York, USA, 2009, pp. 245–248.

[133] L. Baltrunas, M. Kaminskas, et al., Incarmusic: Context-aware music recommendations in a car, In *EC-Web 2011, Toulouse*, Springer, 2011, pp. 89–100.

[134] M. Braunhofer, M. Kaminskas, et al., Location-aware music recommendation, *International Journal of Multimedia Information Retrieval* **2**(1) (2013), 31–44.

[135] L. Baltrunas, T. Makcinskas, et al., Group recommendations with rank aggregation and collaborative filtering, In *RecSys '10* (2010), 119–126.

[136] P. Cremonesi, Y. Koren, et al., Performance of recommender algorithms on top-n recommendation tasks, In *RecSys '10*, ACM, New York, NY, USA, 2010, pp. 39–46.

[137] P. Cremonesi, F. Garzotto, et al., Investigating the persuasion potential of recommender systems from a quality perspective: An empirical study, *ACM T. Interact. Intell. Syst.* **2**(2) (2012), 1–41.

[138] P. Cremonesi, F. Epifania, et al., User profiling vs. accuracy in recommender system user experience, In *AVI '12*, ACM, 2012, pp. 717–720.

[139] M. Elahi, F. Ricci, et al., Active learning strategies for rating elicitation in collaborative filtering: A system-wide perspective, *ACM T. Intel. Syst. Tech.*, to appear, 2014.