MACHINE LEARNING AND DATA MINING

FLORIANA ESPOSITO
Dipartimento di Informatica - Università degli Studi di Bari

ATTILIO GIORDANA, LORENZA SAITTA
Dipartimento di Informatica - Università degli Studi del Piemonte Orientale

1 What Machine Learning is About

Even though learning escapes precise definitions, there is a general agreement on Langley's idea [33] of learning as a set of "mechanisms through which intelligent agents improve their behaviour over time", which seems reasonable once a sufficiently broad view of "agent" is taken. Then, Machine Learning can be roughly described as the process of discovering regularities from a set of available data (examples) and extrapolating these regularities to new data.

Computational studies of learning dates back to the beginning of Artificial Intelligence (AI), with some sparse works in the early 1970's (e.g., [51]). However, the field started to acquire an autonomous profile at the end of the 1970's, when Jaime Carbonell, Ryszard Michalski and Tom Mitchell organized the first workshop devoted to Machine Learning at CMU (Pittsbugh, PA) in 1980. That workshop and the three following ones, in Monticello, IL (1983), at Skytop, PA (1985), and in Irvine, CA (1987), were meant to put up a community, and, hence, they were based on invited attendance, based on a description of research interest. The growing number of people wanting to attend the meeting and the variety of approached themes suggested to launch, in 1988, the first open conference, which was held in Ann Harbor, MI. In the meantime, the *Machine Learning* journal had started to appear in 1986.

For the first decade of its recognized life Machine Learning (ML) flourished mostly in USA, but a few European researchers started, in the early 1980's, to understand the potential of the topic and to work on it. We have to wait until 1992 to see the Conference, which in the meantime became THE forum to publish any substantial Machine Learning progress, held in Europe, in Aberdeen, UK. In parallel, a European conference, the ECML, started to become the reference point for European (but not only) researchers working in the field.

These first years are now viewed as the heroic times in which pioneers, armed with more passion than rigor, entered unexplored regions, where challenging and intriguing phenomena waited to be uncovered.

Now, after more than 25 years, Machine Learning has not only reached full maturity, but it is still growing in number of researchers, in breadth of problems, and in sophistication of methodologies, gradually changing from a subfield of AI to a truly autonomous discipline. Not only this, but Machine Learning has also given birth to a branch that brought it to a new dimension: in 1989, from a workshop organized by Gregory Piatetsky-Shapiro at IJCAI 1989, in Detroit, MI, the Data Mining (DM) and Knowledge Discovery in Data Bases (KDD) originated. The possibility of exploiting Machine Learning techniques to extract "knowledge" from the huge mass of today available data brought these techniques to the attention of an immense set of potential users in all application domains.

The relationships among Machine Learning, Data Mining and Knowledge Discovery in Data Bases did not go without problems. At the beginning, there was a confusion about the coverage of the terms. Now, the received view is that KDD denotes the whole process of extracting knowledge, from data collection and pre-processing to results interpretation. Data Mining is the step, inside this KDD process, in which information is actually extracted from the data by applying algorithms to them. These algorithms are more often than not Machine Learning ones. The KDD field has now its established annual conferences, the ACM KDD and the PKDD (European) conferences, and its own journal, i.e., *Data Mining and Knowledge Discovery*.

The opening of Machine Learning to new horizons makes difficult to trace all the paths it is going down. Then, we will provide, in the next section, only a high level view of the current methodological and applicative landscape.

2 State of the Art

Machine Learning keeps constantly changing, and people attending ICML 2006 (which was held in Pittsburgh, USA, where it all started) could hardly believe to be at the same conference as in Ann Harbor in 1988. Topics ranking high on the research agenda a few years ago, like integration of Machine Learning and Statistics, have already been achieved to some extent and results from Statistics and Pattern Recognition are routinely used, cited and extended in Machine Learning papers.This is one of the reasons that make difficult to depict a clear state of the art of Machine Learning today. Another reason is the adaptation of ML techniques to application fields: methods used in BioInformatics have their own peculiarities, as well as, for instance, the standard learning components embedded in Customer Relationship Management Systems. However, these very peculiarities made them successful in the respective domains.

Nevertheless, in the middle of this diversity, some fundamental issues and trends can be identified, which are methodological and transversal to all application domains.

The first common issue is rigor and awareness in results assessment: with the interaction with real users, results must be carefully and knowingly validated. This need generated a surge of papers on the analysis of commonly used error estimation methods [17], such as cross-validation, and also on the transfer to ML of methods exploited since long in other disciplines, such as the ROC curves [37]. Instead of trying to show that their algorithm is *the best*, researchers try to explore the limits of the algorithm's applicability; a great contribution to this mind change is due to Wolpert's *No-Free-Lunch* theorem [52].

Another general issue is *scaling up* the existing algorithms. In todays applications, where data bases with millions of possibly complex records are to be processed, most classical ML algorithms cannot work. Then, computational requirements become an issue to be addressed. Two lines of attack have been tried: to modify the algorithms in such a way that they can work on very large sets of examples (e.g., [18]), or to reduce the data through a careful pre-processing [46].

On the methodological front, a breakthrough in learning classifiers has been offered by Schapire, who showed that the notions of *strong* learnability and *weak* learnability are equivalent [47]. Since then, ensemble learning has been actively investigated. ¿From the empirical point of view, ensemble learning shows an amazing effectiveness and robustness to overfitting. In the attempt to explain this appealing behavior, much theoretical and empirical work has been done, even though conclusive results are still to be obtained [26, 24].

One of the main sources of requests for new advances is the complex nature of today available data. At its beginning, Machine Learning was applied to numerical/categorical data sets, held in central memory, where examples were represented as vectors of (Attribute - Value) pairs. Now, in an increasing number of domains the interesting data are structured, i.e., each example is composed by interrelated parts, for which this simple representation is no more adequate. Learning from such kind of data is called *Relational learning* [44] or *Inductive Logic Programming* (ILP) [43]. Unfortunately, relational learning suffers from severe (possibly unavoidable) computational problems [29].

Several ways have been tried to reduce the above problem, but without obtaining a definite success. One is *propositionalization*, which is the process of transforming each structured example into a set of unstructured ones, on which the wealth of classical ML algorithms can work [32]. Even though some beneficial effect has been reported, the method is not guaranteed to work, because the number of transformed examples can become exponentially large. Another way is to use a hierarchical approach: examples can be abstracted and moved to another representation space in which learning becomes hopefully simpler; again, if the ”good” abstraction is found, things work out well, but finding this abstraction may be prohibitively hard [53].

A recently proposed approach, which takes into account some relation in the data without facing full relational learning, is *Statistical relational learning* [27]. The approach is based on probabilistic relational models, which extend probabilistic graphical models with the expressive power of object-relational languages. Probabilistic relational models handle the uncertainty over the attributes of objects in the domain as well as uncertainty over the existence of relations between objects These models can be automatically learned from a relational data set, and they extend Bayesian networks learning.

Another approach that is currently at the core of Machine Learning is the *Kernel-based* one. Starting from the seminal paper by Vapnik on Support Vector Machines [16], several improvements and applications have been proposed (see, for instance, [14]).

Reinforcement learning [50] another branch of Machine Learning which is receiving attention, especially for extending it to relational domains [5]. However, the computational cost required still hinders it to be successfully applied in practice, except in very rare cases.

Learning from complex objects, such as textual documents, spatio/temporal data, images, video, or multimedia is currently a major challenge for Machine Learning [15]. Whereas indexing and categorization of textual documents have already received much attention and good results have been obtained [31], learning to categorize or retrieve images by their content is still at the beginning. Object categorization, in fact, is difficult because differing pose, scale, illumination and intrinsic visual differences produce highly different images for objects of the same class. Existing shape-based modeling techniques are not designed to deal with these large variations. The challenge in object categorization is to find class models that are invariant enough to incorporate naturally occurring intra-class variations and yet discriminative enough to distinguish between different classes.

In the media and entertainment industries, including streaming audio and digital TV, generated data present new challenges for Machine Learning [35]. Regarding automatic video indexing, for instance, current content-based systems use low-level features, such as motion, color, and texture. However, low-level features often have little meaning for users, who much prefer to identify content using high-level semantics or concepts. This creates a gap between systems and their users that must be bridged for these systems to be able to mining semantically meaningful information.

Spatio-temporal data mining is an emerging research area dedicated to the development and application of novel computational techniques for the analysis of large spatio-temporal databases [2]. Both the temporal and spatial dimensions add substantial complexity to data mining tasks. First of all, the spatial relations, both metric (such as distance) and non-metric (such as topology, direction, shape, etc.) and the temporal relations (such as before and after) are information that needs to be considered in the data mining methods. Moreover, some spatial and temporal relations are not explicitly encoded in a database, and must be extracted from the data. Finally, working at the level of stored data (points, lines, time stamps) is often undesirable. Therefore, complex transformations are required to describe the units of analysis at higher conceptual levels, where human-interpretable properties and relations are expressed. Finally, spatial resolution or temporal granularity have direct impact on the strength of patterns that can be discovered. All the above reasons make the integration of spatio-temporal reasoning in data mining still an open problem.

An issue that is more and more debated both inside and outside the Knowledge Discovery/Machine Learning community is *privacy* and *security* (e.g., [34, 42]. Data Mining, with its need to access all kinds of data, arises fears of privacy violation. Although techniques, such as random perturbation techniques, secure multi-party computation based approaches, cryptographic-based methods, and database inference control have been developed, many of the key problems still remain open in this area and they will affect the design of data mining and learning algorithms. The security and privacy issues are the more important now, as government agencies start to extensively use knowledge discovery tools, and mining official data is becoming a hot topic.

2.1 The Italian landscape

Machine Learning, in its traditional meaning, has been present in Italy since the beginning of the 1980's, when researchers at the Universities of Torino and Bari started to work on the subject. This group of academic pioneers was joined by researchers from the *Ugo Bordoni* Foundation and *IBM Italia*, in Rome, by people at *CSELT* (now TiLab) in Torino, as well as by researchers from other universities. The Italian community reached quickly a good international visibility, and was present, through its members, at all important moments of the international life, notably, *Machine Learning*'s and *Journal of Machine Learning Research*'s editorial boards, ICML's, ECML's, KDD's and PKDD's programs committees, and also ICML's and ECML's Program Chairpersonships.

The core of the original groups of people remained faithful to Machine Learning (even when part of the Torino team moved to Alessandria, campus of the new University of Piemonte Orientale). Both teams in Torino/Alessandria and Bari grew in size and in breadth of themes, as several Ph.D. students were attracted by the challenges of the field.

In the meantime, the community expanded with the addition of groups at the Universities of Pisa, Roma ”La Sapienza”, Milano, Firenze and Padova, as well as from IRST and companies such as *TiLab* and *Alenia Spazio* (both in Torino).

From the scientific point of view, the research topics in Machine Learning concentrated first on concept learning for classification tasks, including the whole spectrum of supervised approaches, from decision tree [21] and model tree [40] building to relational learning ([7], [19], [38]), from genetic algorithms [3] to neural networks (e.g., [1, 8]) and kernel machines ([48], see also the parallel article in this issue), and unsupervised approaches (clustering) (e.g., [6, 11, 49, 30]). Also researches in Computational Learning theory (e.g., [13, 4]), complexity of learning [10], changes of representation [45], [25], and incremental methods [22], [20] are very active. More on the Data Mining side, the focus is on extracting association rules from databases [12, 23, 28, 36], handling geographical information systems [41], [39] and automatically building models of sequences ([9]).

3 Open Problems and Research Perspectives

With the diffusion of a new society, heavily based on networking and knowledge acquisition and communication, a new concept has been forged: users are no longer obliged to do all efforts to understand and master their environment, but rather they interact with an ”ambient intelligence”, aiming at ”understanding” them, seamlessly supporting their activities, and engaging the environment in adaptive cooperation on a personalized basis. The future Learning Science can provide a transparent mediation between the user and the data in this new context.

Recently, topics such as distributed systems, complex and heterogeneous data and process integration have become important topics on the agenda. We are already surrounded by devices that embed microprocessors, such as consumer electronics, medical devices, cars, and so on, but they generally operate stand-alone. In the ubiquitous computing vision, all these devices are not only capable of computing but also of communicating: each networked device may take advantage of the services offered by other connected devices instead of duplicating their functionality. Ubiquitous computing represents a paradigm shift in the history of information and communication technologies, according to which computing is omnipresent and devices that do not look like computers are endowed with computing capabilities.

In order to achieve this goal, future researches should address at least the following topics:

- *knowledge discovery in mobile systems*: mobile systems, wireless communication networks, calm technologies,
- *distributed architectures*: distributed data mining, grid computing, global computing, peer-to-peer, autonomic computing, agent technology and embedded data mining,
- *learning components*: any-time learning, graph-based and relational learning, statistical learning, evolutionary computation,
- *data types*: spatio-temporal data, stream data, audio,-video, text data, web data,
- *security and privacy*: privacy preserving data mining, intrusion detection
- *human computer interaction*: user interfaces of ubiquitous discovery systems, visual knowledge discovery.

The first three items constitute a paradigm shift for the field of knowledge discovery, since the idea of a standalone (desktop or workstation) analysis tool is abandoned in favour of process-integrated, distributed and autonomous analysis engines. This technical approach can only succeed if the topics of privacy and security are addressed in a much more principled and multi-disciplinary manner than before.

A further major objective is to re-investigate the relation between machine and human learning. The shift to mobile, distributed and embedded applications utilizing heterogeneous data sources is paralleled by a shift in the cognitive sciences towards embodiment. According to this view, cognition is structured by interaction with the environment and grounding of concepts in (combinations of) sensory inputs. For this reason, the agenda of knowledge discovery and machine learning is now re-oriented to fully account for the user in the loop, and investigate in a principled way the development of data mining systems with cognitive abilities.

Finally, it is responsibility of data mining researchers to invent theories, methods and techniques, which are aware of citizens rights of privacy and confidentiality. A privacy-aware technology that is designed to prevent infringing civil rights would enable a wider social acceptance of a multitude of new services and applications that would find in ubiquitous knowledge discovery a key driver.

Fortunately, the field of machine Learning and Knowledge Discovery is ready for the transformation. This requires to take up the most advanced research results in various sub-areas and to carefully integrate them. This combination of technologies will lead the field far beyond the current the state of the art. We can sum up the challenges if we formulate the problems from a data-centric perspective. Compared to more traditional environments the challenges are created by the following key features of ubiquitous environments:

- the amount of data, which makes impossible manual approaches or exhaustive mechanisms for screening, filtering, aggregating and managing data,
- ad-hoc networks, which do not allow strict assumptions on the trustworthiness and/or technical reliability of the network nodes,
- heterogeneous data, which do not allow monolithic representation schemes,
- distributed data, which complicate the process of finding relevant information and managing it,
- continuous data streams, which forbid one-time ad hoc solutions,
- real-time data, which put high demand on efficiency and run-time complexity.

Principled, theoretically well-founded approaches for addressing these problems will soon become mission critical. No substantial progress can be made without addressing the problem of learning from data. More specifically, the goal is to build systems with greater autonomy, flexibility, intelligence that are:

- self describing, self-organizing,
- self-monitoring and repairing, and can perform automatic intrusion detection,
- perform automated trend detection, prediction, aggregation and summarization,
- are inherently distributed.

4 Links with other AI fields

Even though Machine Learning was considered, at its start, as a subfield of Artificial Intelligence, over the years it became apparent that the two fields intersect, but neither one includes the other. In fact, Machine Learning receives conceptual and methodological contributions from several disciplines outside AI, as described in the next section. Nevertheless, AI strongly motivates ML, as the ability to learn is a hallmark of intelligence, and heavily contributes with methodologies as well.

More specifically, one of the subfield ML draws upon is *knowledge representation*. The choice of the languages for describing data and hypotheses in any ML system has been recognized since ever as a critical one, whatever the approach the system is based upon. As soon as Machine Learning started targeting learning knowledge comprehensible by humans, logical languages came spontaneously into play. At the beginning, simple, propositional languages were mostly used, to cope with the naive algorithms available at the time and the undemanding applications. Later on, more complex logical languages, such as Horn clauses, Datalog and Description logic have been used, exploiting the wealth of results about representation power and inference ability. Recently, more sophisticated issues, such as abstraction and automatic changes of representation, suggested ML new ways to cope with computational complexity.

Another fundamental AI subfield, without which ML would probably not exist, is *search*. Since Tom Mitchell recognized, in 1980, that learning is exactly a search in the space of possible hypotheses, results in this AI's core subfield moved steadily to ML, for the good (continuously improving search techniques) and for the bad (a phase transition in the task of matching hypotheses and examples emerged).

The above two subfields have been sources of essential contributes to the building of ML as an autonomous discipline. Many others have with ML a different type of relation, i.e., they provide problems to be solved (learning in planning, scheduling, diagnosis, natural language processing, vision, and so on), and, hence, they are users of ML approaches.

With the birth of Data Mining, a preheminent place has been taken by Data Bases. While early ML systems worked with data kept in main memory, the size of todays real-world data warehouses requires that Data Mining tools be applied directly to the data bases. Also, new methodologies, such as association rules extraction, have received a lot of interest from researchers working on data bases.

5 Links with other Disciplines

Machine Learning, with its broad coverage of subjects, is a highly interdisciplinary area, encompassing Statistics, Pattern Recognition and Signal Processing, Control theory, Information and Communication theory, Cognitive Sciences, Evolutionary theories, Neurobiology, and even Philosophy. Recently, it established links also with distributed, grid and mobile computing.

Statistics offers tool for data analysis since more than a century, but Machine Learning, targeting (at least in the idea of its initiators) a type of learning closer to human conceptual knowledge acquisition, started as an alternative to Statistics, considered to be too low level to be interesting. Only time showed that ML cannot avoid Statistics as a fundamental tool for evaluating its results, and that some statistical methods can be fruitfully integrated into more symbolic approaches. When Data Mining entered the scene, again Statistics was perceived as a competitor, but the two can readily coexist and be synergic: Statistics follows a *veriÞcation-driven* philosophy, where a human expert formulates a model, which is then checked against the data; in this way, past experience can be fully exploited, but it is difficult to find something really new. On the contrary, Data Mining follow a *discovery-driven* approach, where data ”are let speak by themselves”; in this way, a-priori knowledge does not bias the discovery process, but the search is more difficult.

Pattern Recognition and Signal Processing share with Machine Learning many tasks, for instance classification, recognition, and clustering, and in some cases also the methods (for example, building decision tree originated in Pattern Recognition, but this was not acknowledged by early ML researchers). It is not easy to tell which is the real difference (if any) between Pattern Recognition and Machine Learning, especially todays, when algorithms used in Pattern Recognition have become more sophisticated and similar to those used in ML, whereas ML extended its cope to include tasks, such as image classification, which are typical Pattern Recognition's ones. If a difference has to be found, it may reside in the Pattern Recognition preference for numerical approaches dealing with low level (close to the signal) features, whereas Machine Learning has still the ambition to look like human conceptual learning. It is clear, in any case, that many ML tasks, such as learning in vision, require the results previously obtained by Pattern Recognition.

Information theory is largely used to guide the search in the space of hypotheses. Many notions borrowed from it, such as entropy, Gini index, Information gain, and several others are used as heuristics during learning. Another notion, linked to both Information and Communication theories, i.e., the *Minimum Description Length* is a largely used search heuristics. Finally, the ideas of *Error Correcting Code* and *Vector quantization* underly a classification and a clustering approach, respectively.

A branch of Machine Learning, namely *Reinforcement Learning*, can be applied both to learning how to control dynamic systems, and to learning how to explore, discover and act in an environment. This approach, inspired by the *stimulus-response* learning in biological systems, is not the unique to have this origin. In fact, the Darwinian theory of evolution has been adapted to the Evolutionary Computation approach (encompassing Genetic algorithms, Tabu search and Simulated annealing), whereas the organic brain is questionably imitated by neural nets.

It goes without saying that Cognitive Sciences have very many contacts with Machine Learning. Concept formation corresponds to clustering, learning algorithms have been used to model aspects of human learning or to describe conceptual change, human perception is a very rich sources of suggestions for automatic processing (for instance in vision), and student modelling has been tested on young children in school. Even though some cooperation between Machine Learning and Cognitive Sciences is already a reality, most of the work is still left to the future.

Finally, Machine Learning touches issues that are in the philosophical realm: for instance, the very notion of what learning is, the relation between induction and abduction, the relation between analogy and similarity, the very idea of data compression and minimum description length (or Kolmogorov complexity), and many others.

6 Application Achievements

In the last decade Machine Learning left the protected environment of academy and entered the real world. After a slow start, good results began to convince the users of the worthiness of the methodology, and now it is impossible to mention all the successes Machine Learning has achieved. This is especially true if we also include Data Mining, which is by now entering most of the big economic, industrial and governmental entities. In 2002 started the "International Conference on Machine Learning and Applications" (ICMLA), which was in 2005 at its 4th edition: the conference is a forum for advanced ML applications in all domains.

An overview of Machine Learning's achievements is better organized according to the application field. A major one, coming immediately to mind, is Medicine, which was one of first on which ML methods have been tried. Medical Informatics can play a key role in improving quality and performance of health care, and thereby improving health status of the population. To enable clinicians to improve their care information is needed.

Evidence based medicine has developed as the guiding principle in everyday health care. Clinical trials are performed on a large scale, structured reviews present data in useful ways, evidence based clinical practice guidelines summarize available knowledge for health care workers. Yet, this knowledge is not necessarily available to the busy clinician, at the moment she makes a decision. Knowledge Discovery can summarise and tailor the data such that they fit the clinicians needs. Also, it is certain that not all components of medical practice will be subjected to prospective randomized trials, clinical reality may be different from the optimal care as suggested by these trials and trials usually include 'healthier' patients than patients consulting clinicians in daily practice. Taking all this together, there is a clear need for information about the reality of clinical practice.

Molecular Biology followed: DNA and proteins analysis owes a lot to Machine Learning, which contributed with classification and clustering algorithms and with sequence-oriented methods (such as Hidden Markov Models) to the location of genes and regulatory patterns, and to protein structure identification (Workshops at ICML 2003, KDD 2003, and PKDD 2004). Currently much work is devoted to micro-arrays analysis. Application of discovery and learning tools to micro-arrays data, commonly called "gene chips", made it possible to simultaneously measure the rate at which a cell or tissue is expressing each of its thousands of genes. One can use these comprehensive snapshots of biological activity to infer regulatory pathways in cells, identify novel targets for drug design, and improve the diagnosis, prognosis, and treatment planning for those suffering from disease.

The Web, this immense information repository, has become a privileged target of Machine Learning: classification, clustering and indexing of (textual) documents can be done automatically now, thanks to ML techniques. For instance, text classification via Support Vector Machines (SVM) reached impressive performances. A new ranking SVM has been developed which exploits click-through data on the web for personalisation. Also to Information Retrieval and Information Extraction Machine Learning approaches proved to be fundamental. A common task, in this context, is user's profiling, performed by e-commerce companies, recommender systems, or network intrusion detector systems.

Automatic profiling (which exploits statistical user characterization or more sophisticated approaches, such as Hidden Markov Models) is also popular in banking and insurance, where a major problem is risk assessment, and in telecommunication and credit card fraud detection, where anomalous user behaviours must be detected. User behaviour consists of typical action sequences performed by people. Learning techniques are exploited to automatically synthesize action sequences from log files; then, the individual user behaviour is matched against such sequences to understand what s/he is doing and possibly help her/him at the execution of the task.

Language learning is also an active area of research that uses language models. Language models, which provide the probability of word sequences, are used in speech recognition, machine translation, and many other areas. Current techniques in language modelling include word clustering and smoothing (regularization), and also more language-model specific techniques such as high order n-grams and sentence mixture models.

A major role played by Machine Learning and Data Mining in companies is that of a supportive (for the effective retrieval, characterisation and identification of existing knowledge artefacts) and enabling technology (for the extraction of knowledge from datasets and information artefacts).

Computational science has historically meant simulation; but, there is an increasing role for analysis and mining of online scientific data. As a case in point, half of the world's astronomy data is public. The astronomy community is putting all that data on the Internet so that the Internet becomes the world's best telescope: it has the whole sky, in many bands, and in detail as good as the best 2-year-old telescopes. It is useable by all astronomers everywhere. This is the vision of the virtual observatory, also called the World Wide Telescope (WWT). Microsoft has invested heavily in this project (Gray, KDD 2003).

In statistics, the term official data denotes data collected in censuses and statistical surveys by National Statistics Institutes (NSIs), as well as administrative and registration records collected by government departments and local authorities. A special issue of the journal Intelligent Data Analysis appeared in December 2003 offers the state of the art in the domain).

7 Application Potential

The Information Society Technologies Advisory Group (ISTAG) has recently identified a set of grand research challenges for the UE 7th Framework FP7. Among these challenges are the totally safe car, a multilingual companion, a service robot companion, the self-monitoring and self repairing computer, the Internet police agent, a disease and treatment simulator, an augmented personal memory, a pervasive communication jacket, a personal everywhere visualiser, an ultra light aerial transportation agent, and the intelligent retail store.

By analysing these challenges, it becomes clear that most of them require systems that are able to adapt to their environment, to learn from past experience, and to improve their performance by learning from past mistakes. This learning component shows up in the frequent use of terms such as intelligent, smart, adaptive in these scenarios. Then, building systems that learn is an underlying common thread in these grand challenges.

Knowledge Discovery has always been directed towards technology developments and commercial applications. The time elapsing from a theoretical breakthrough to commercial application is often significantly less then 10 years. The financial barriers for entrance of SMEs and start-up companies in this high technology market area are lower than in other markets. Its broad scope makes it an enabling technology for the knowledge society and e-Europe, thus answering to a genuine demand. Both factors (application orientation and market need) combine in making Knowledge Discovery a research area with among the highest potentials for implementing structures that ensure rapid transfer of knowledge from research to industry. This in turn will create a favourable environment for firm creation. So even if research in this new view of knowledge discovery is currently in a very early phase, it is important to set the stage for tomorrows commercial breakthroughs right now.

REFERENCES

[1] F. Aiolli and A. Sperduti. Multiclass classification with multi-prototype support vector machines. *Journal of Machine Learning Research*, 6:817–850, 2005.

[2] G. Andrienko, D. Malerba, M. May, and M. Teisseire, editors. *Proceedings of ECML/PKDD Workshop on Mining Spatio-Temporal Data*, Porto, Portugal, 2005.

[3] C. Anglano, A. Giordana, G. L. Bello, and L. Saitta. An experimental evaluation of coevolutive concept learning. In *Proc. 15th International Conf. on Machine Learning*, pages 19–27. Morgan Kaufmann, San Francisco, CA, 1998.

[4] B. Apolloni, D. Malchiodi, and S. Gaito. Algorithmic inference in machine learning. In *Advanced Knowledge International*, volume 5 of *International Series on Advanced Intelligence*. 2003.

[5] A. Barto, C. Boutilier, K. Driessens, S. Dzeroski, R. Givan, C. Guestrin, S. Russell, and P. Tadepalli, editors. *Proc. ICML Workshop on Relational Reinforcement Learning*, Banff, Canada, 2004.

[6] R. Basili, M.-T. Pazienza, and P. Velardi. A context driven conceptual clustering method for verb classification. In *Corpus processing for lexical acquisition*, pages 117–142. MIT Press, Cambridge, MA, USA, 1996.

[7] F. Bergadano, A. Giordana, and L. Saitta. *Machine Learning: A General Framework and its Applications*. Ellis Horwood. Chichester, UK, 1991.

[8] M. Bianchini, M. Gori, L. Sarti, and F. Scarselli. Recursive processing of cyclic graphs. *IEEE Transactions on Neural Networks*, 17(1):10–18, 2006.

[9] M. Botta, U. Galassi, and A. Giordana. Learning complex and sparse events in long sequences. In *Proc. of the 16th European Conference on Artiﬁcial Intelligence*, pages 425–429, 2004.

[10] M. Botta, A. Giordana, L. Saitta, and M. Sebag. Relational learning as search in a critical region. *J. Machine Learning Research*, 4:431–463, 2003.

[11] C. Carpineto and G. Romano. A lattice conceptual clustering system and its application to browsing retrieval. *Machine Learning*, 24(2):95–122, 1996.

[12] C. Carpineto and G. Romano. Mining short-rule covers in relational databases. *Computational Intelligence*, 19(3):215–234, 2003.

[13] N. Cesa-Bianchi and G. Lugosi. Potential-based algorithms in on-line prediction and game theory. *Machine Learning*, 51(3):239–261, 2003.

[14] W. Cohen and A. Moore, editors. *Proceedings of the International Conference on Machine Learning (ICMLÕ06)*. Morgan Kauffman, San Francisco, CA, 2006.

[15] M. Cord, P. Cunningham, T. Sziranyi, and R. Dahyot, editors. *Proc. ICML Workshop on Machine Learning Techniques for Processing Multimedia Content*, Bonn, Germany, 2005.

[16] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[17] T. Dietterich. Approximate statistical test for comparing supervised classification learning algorithms. *Neural Computation*, 10:1895–1923, 1998.

[18] P. Domingos and G. Hulten. A general method for scaling up machine learning algorithms and its application to clustering. In *ICML Õ01: Proceedings of the Eighteenth International Conference on Machine Learning*, pages 106–113, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[19] F. Esposito, N. Fanizzi, S. Ferilli, and G. Semeraro. A generalization model based on oi-implication for ideal theory refinement. *Fundamenta Informaticae*, 47(1-2):15–33, 2001.

[20] F. Esposito, S. Ferilli, N. Fanizzi, T. M. A. Basile, and N. D. Mauro. Incremental learning and concept drift in INTHELEX. *Intelligent Data Analysis*, 8(3):213–237, 2004.

[21] F. Esposito, D. Malerba, and G. Semeraro. A comparative analysis of methods for pruning decision trees. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 19(5):476–491, 1997.

[22] F. Esposito, G. Semeraro, N. Fanizzi, and S. Ferilli. Multistrategy theory revision: Induction and abduction in INTHELEX. *Machine Learning*, 38(1-2), 2000.

[23] R. Esposito, R. Meo, and M. Botta. Answering constraint-based mining queries on itemsets using previous materialized results. *J. Intell. Inf. Syst.*, 26(1):95–111, 2006.

[24] R. Esposito and L. Saitta. A Monte Carlo analysis of ensemble classification. In *ICML Õ04: Proceedings of the twenty-Þrst international conference on Machine Learning*, page 34, New York, NY, USA, 2004. ACM Press.

[25] S. Ferilli, T. M. A. Basile, N. D. Mauro, and F. Esposito. On the learnability of abstraction theories from observations for relational learning. In *Machine Learning: ECML 2005, 16th European Conference on Machine Learning, Proceedings*, volume 3720 of *Lecture Notes in Computer Science*, pages 120–132. Springer, 2005.

[26] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting (with discussion). *Annals of Statistics*, 28(2):337–407, 2000.

[27] L. Getoor, N. Friedman, B. Taskar, and D. Koller. Learning probabilistic models of relational structure. In *Journal of Machine Learning Research*, volume 3, pages 679–707, Hingham, MA, USA, 2002. Kluwer Academic Publishers.

[28] F. Giannotti, M. Nanni, D. Pedreschi, and F. Pinelli. Mining sequences with temporal annotations. In *Proc. ACM SAC*, pages 593–597, 2006.

[29] A. Giordana and L. Saitta. Phase transitions in relational learning. *Machine Learning*, 41(2):217–251, 2000.

[30] T. Hofmann and J. Buhmann. Inferring hierarchical clustering structures by deterministic annealing. In *Proc. of KDD*, pages 363–366, 1996.

[31] T. Joachims. *Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA, 2002.

[32] M.-A. Krogel, S. Rawles, F. Zelezný, P. A. Flach, N. Lavrac, and S. Wrobel. Comparative evaluation of approaches to propositionalization. In *Proc. ILP*, pages 197–214, 2003.

[33] P. Langley. The science of machine learning. Preface. In *Proc. of the Seventeenth International Conference on Machine Learning (ICML-2000)*. Morgan Kaufmann, 2000.

[34] S. Laur, H. Lipmaa, and T. MielikŁinen. Cryptographically private support vector machines. In *Proc. of 12th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining*, Philadelphia, USA, 2006.

[35] W. K. Leow, M. Lew, T. Chua, W. Ma, L. Chaisorn, and E. Bakker, editors. *Proc. 4th Int. Conf on Image and Video Retrieval*, volume 3568 of *Lecture Notes in Computer Science*. Springer, 2005.

[36] F. A. Lisi and D. Malerba. Inducing multi-level association rules from multiple relations. *Machine Learning*, 55(2):175–210, 2004.

[37] S. Macskassy, F. Provost, and S. Rosset. Roc confidence bands: an empirical evaluation. In *ICML*, pages 537–544, 2005.

[38] D. Malerba. Learning recursive theories in the normal ILP setting. *Fundamenta Informaticae*, 57(1):39–77, 2003.

[39] D. Malerba, A. Appice, and M. Ceci. A data mining query language for knowledge discovery in a geographical information system. In R. Meo, P. L. Lanzi, and M. Klemettinen, editors, *Database Support for Data Mining Applications*, pages 95–116, 2004.

[40] D. Malerba, F. Esposito, M. Ceci, and A. Appice. Top-down induction of model trees with regression and splitting nodes. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 26(5):612–625, 2004.

[41] P. Mancarella, A. Raffaetà, C. Renso, and F. Turini. Integrating knowledge representation and reasoning in geographical information systems. *International Journal of Geographical Information Science*, 18(4):417–447, 2004.

[42] S. Matwin, L. Chang, R. Wrtight, and J. Zhan, editors. *Proc. of IEEE Workshop on Privacy and Security Aspects of Data Mining*, Houston, Texts, USA, 2005.

[43] S. Muggleton. *Inductive logic programming*. Academic Press, 1992.

[44] J. Quinlan. Learning logical definitions from relations. *Machine Learning*, 5(3):239–266, 1990.

[45] L. Saitta and J.-D. Zucker. A model of abstraction in visual perception. *Applied Artiþcial Intelligence*, 15(8):761–776, 2001.

[46] C. Saunders, M. Grobelnik, S. Gunn, and J. Shawe-Taylor, editors. *Subspace, Latent Structure and Feature Selection, Statistical and Optimization, Perspectives Workshop*, volume 3940 of *Lecture Notes in Computer Science*. Springer, 2006.

[47] R. Schapire. The strength of weak learnability. *Machine Learning*, 5(2):197–227, 1990.

[48] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[49] D. Sona, S. Veeramachaneni, P. Avesani, and N. Polettini. Clustering with propagation for hierarchical document classification. In M. Gori, M. Ceci, and M. Nanni, editors, *ECML/PKDD Proceedings of the Workshop on Statistical Approaches for Web Mining*, pages 50–61, 2004.

[50] R. Sutton and A. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.

[51] P. Winston. Learning structural descriptions from examples. In P. Winston, editor, *The psychology of computer vision*, pages 157–209, New York, NY, 1975. McGraw Hill.

[52] D. Wolpert and W. Macready. No free lunch theorems for search (technical report). Technical Report SFI-TR-95-02-010, Santa Fe Institute, 1995.

[53] J. Zucker and L. Saitta, editors. *Proc. 6th Intern. Symposium on Abstraction, Reformulation and Approximation*, volume 3607 of *Lecture Notes in Computer Science*. Springer, 2005.