

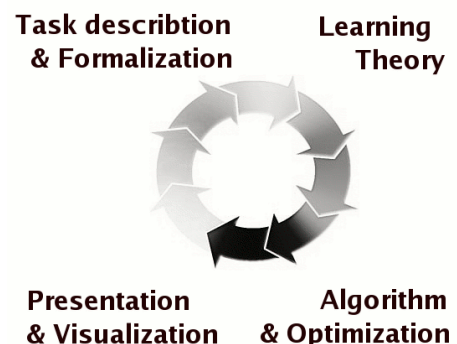
Design of Learning Machines, Algorithms and Analysis

Research Summary of Dr. Kristiaan Pelckmans

Techniques for learning from observational data play a critical role in countless branches of science nowadays. Although monopolized by the field of statistics for many decades, the modern area of Machine Learning (ML) complements the more and more these classical techniques, thereby emphasizing a number of needs which are more intended for *predictive* and *exploratory* purposes in a *complex and structured environment*. A powerful feature of the area is the importance of cross-relating a wide range of scientific breakthroughs, as such providing a framework for *interdisciplinary* research. Thirdly, the concern with *computational, algorithmic and optimization* issues personalize the machine learning field. The main personal motivation for being involved/engulfed by the field is the potential it has for systematically discovering hidden structures, and automatically exploring this complex world. This view is reinforced by the inspiration and creativity as apparent in works by amongst many others L. Devroye, and S. Boyd et al.. My research on ML is based from study on the crossroads between such different fields as convex optimization, system identification, nonlinear systems, signal processing, time-series analysis, parametric and non-parametric statistics, information theory, combinatorial optimization and operational research, and experience in a wide range of application areas.

I. Overall Objective - Identifying and Studying Novel Learning Tasks

My research is characterized by the quest for assembling and analyzing a *methodological toolbox* for designing learning machines for a broad range of tasks. The first major step in the design cycle of a learning machine occurs in the description of the task through communication with the expert user. During subsequent formalization, a major concern is to ensure sufficient covering by theoretical analysis as described e.g. in statistical learning theory. In general, the literature on statistical learning theory provides necessary and sufficient conditions for a machine to learn effectively from data, as such guaranteeing scientific validity of the results, and feasibility of the task at hand.



Historically, theoretical studies resulted in various principled approaches for motivating a specific learning machine, with main players the classical VC and PAC-style results, PAC/Bayesian analysis, and mistake bound methods and theoretical results on specific learning machines as the perceptron algorithm. Especially the compression approach - linking directly to information theory and nonparametric statistics - is conceived as a very flexible and powerful theoretical tool, also appealing to practice. The notion of learning is inherently connected to a principle of optimality translated in an actual and empirical *risk term*. Furthermore (and from the start on) the role of restricting the hypothesis space - also denoted as *regularization* - was placed in the very center of the analysis, hereby discriminating somewhat from the field of classical parametric statistics. The design of an application-dependent learning machine requires as such a suitable formalization of a notion of risk or loss, a concept of regularization and of a predictive model, all of which are coherent with the expert knowledge. The methodological framework of primal-dual kernel machines provides a unifying framework for the construction of a wide range of nonlinear predictive models, choosing resolutely an (engineering) convex optimization view. Particularly, the primal-dual derivations are found to be a principled tool for deriving non-trivial machine formulations, providing new insights in the task at hand and complementing the variational approach. Given the resulting algorithm, a final step in the design process is the presentation and the visualization of the result, and eventual re-iterating over the design circle.

Specifically, the interplay between *predictive modeling*, and *un- (or semi-) supervised exploratory tasks* are seen to provide a fertile ground for formulating novel learning tasks. This wide landscape is still sparsely populated with common techniques for e.g. clustering purposes. Through the study of learning algorithms for specific tasks, I hope to find a taxonomy, and to uncover hidden task descriptions in this field of data-mining. Different related elements are found in the following issues

- Clustering and optimal visualization (Practical major approach for getting insight in the data)
- Predictive mutual clustering (Prediction of cluster content)
- Density estimation and outlier detection (Predictions and probabilities)
- Predictive Tolerance Intervals (Qualitative predictions)
- Nonparametric Ordinal Regression (Nonlinear Utility Functions)
- Survival and reliability analysis (Modeling Failure Times)
- Dynamic state estimation and state reconstruction (Subspace identification and linear algebra)

In particular, emerging scientific questions in the context of bio-informatics (e.g. prioritization of genes) and information retrieval (e.g. searching in huge libraries of e-books), and the increasing automatization in medical application areas (e.g. in ICU support) constitute a stimulating force for exploring various subtle manifestations of learning tasks in complex data-rich environments.

This research direction is directly stimulated through involvement with various application areas, see section IV. This research is also aided through interaction with the statistical department, through the FWO project on robust statistics (with M. Hubert, Math, K.U.Leuven), and with colleague Dr. J. De Brabanter. Research on nonlinear system identification was supported by the FWO project on nonlinear system identification together with prof. J. Schoukens (V.U.B, Brussels).

II. Tools - Convex Optimization for the Design of Learning Machines

My *doctoral dissertation* describes a generic and methodological approach for the design of learning algorithms based on techniques of convex optimization, which is entirely in accordance with the engineering background of the research group. The main objective here was to create a learning machine suitable for a specific (application) task, while ensuring a sufficient covering by theoretical results in statistical learning theory, and most importantly, yielding a formal problem statement which can be solved efficiently using a standard convex optimization solver. Recently, techniques of Convex Optimization (CO) take a prominent place in learning approaches, as pioneered by the work on Support Vector Machines (SVMs) and other regularization based learning schemes as regularization networks and Least Squares SVMs (LS-SVMs). Duality theory has played an important role in the development of so-called kernel machines, while the fact of uniqueness of the optimal solution has permitted theoretical as well as practical breakthroughs. A third main advantage of using CO tools in research on learning problems is that the conceptual level of the design of a learning scheme becomes nicely separated from the actual algorithm implementing this scheme (the CO solver). This places the machine learner in the convenient position of interfacing the practical task at hand to the actual algorithm, emphasizing as such the proper defining and tuning of suitable principles and mechanisms for the learning task at hand.

Critical elements as outlined in the previous section reoccur: (1) which criterion of optimality is suitable (e.g. a notion of risk, risk scores, likelihood, or a test statistic), (2) which notion of regularization is appropriate (e.g. a 2-norm, 1-norm, ∞ -norm restriction on the unknowns of the problem, or a combination as e.g. clustering shrinkage) and (3) what further a-priori knowledge is available (e.g. structural constraints as monotonicity, stability requirements or parametric parts). Related especially to the first topic is the analysis of the reason why the model is to be used (is it e.g. purely predictive or also intended to infer some insight on the problem at hand). The study of this toolbox of techniques, and the way they can be encoded as a convex optimization problem forms the main theme of my dissertation. This work was supported by the implementation of the Matlab/C toolbox LS-SVMlab dedicated to LS-SVMs and related algorithms as kernel PCA and Bayesian inference techniques. This work resulted in a reliable overall feel of the critical parts of machine learning techniques in practice, especially through the user-feedback and the hands-on experience observed in many student labs. In

particular, the demand for a reliable procedure for model selection and handling large datasets was taken into account.

During past research, we devoted particular attention to the role of *optimization in model selection* tasks in kernel machines. This emphasis stems from the somewhat unconventional view of the author that the quest for finding structural properties and stable patterns in complex data is essentially a model selection task (as e.g. input selection). Particular attention was given to such critical tasks as (1) selecting the regularization trade-off constant in smoothing techniques as ridge-regression, regularization networks, smoothing splines and LS- SVMs, (2) selecting an appropriate kernel from data, and recently (3) to the task of input selection in a variety of learning algorithms.

This research is stimulated by various research networks: at first, it is entirely integrated in the 'center of excellence: optimization in engineering', established together with a variety of other departments in the KULeuven; additional links are established (IAP5) with the CORE research group of UCL, Louvain-La-Neuve; and a numerical focus was reinforced through the background and the international significance of the research group, and its relation to various other well-known researchers (amongst many others G. Golub and P. Van Dooren).

III. Theory - Learning and Graph Labeling for Empirical Weighted Graphs

My *recent investigations* concentrate on the study of learning problems and algorithms based on empirical (observed) graphs. While many learning problems can be naturally embedded in a proper metric space, a wide range of problems are presented originally as a graph problem. Although my past research has been characterized by an engineering view on the learning point, this research essentially pursues a novel theoretical foundation and the corresponding computational grounds for designing algorithms in this context. It is especially interesting that the main themes - induction, clustering, and complexity - and the main theoretical tools - VC dimension, compression bounds, ... - reappear in this context in a subtly different way than classical, yielding renewed insights in the very essence of learning. The convenient consequence being that this allows the author to be technical, while yielding crisp interpretations and unconventional views (*technical, without the too*).

This investigations aim at *bridging a gap* between research in combinatorial optimization, essentially operating in a discrete universe, and the recent advance in statistical learning and inference methods. The aim is to form an interface between the study of combinatorial algorithms - such as MINCUT-MAXFLOW and minimum spanning trees - to a context of machine learning. This is especially interesting since statistical analysis can potentially provide a different view on complexity issues of algorithms. While the classical view is basically a worst-case approach (focusing on '*P or NP*'), the analysis from a learning point of view complements this with a principled approach to formalize average properties of the resulting estimate. Particularly, the analysis using compression arguments, and stability of learning algorithms operates on the border between those two views. In future work, I plan to explore this relationship further. Especially the application of statistical learning towards searching (in complex unstructured universes) is promising. Another advantage for aiming for dedicated algorithms (complementing the generic CO solvers) for a specific task is the ability to formulate and extend results for huge databases.

A particular interesting *future research direction* targets the integration of learning and inference with the emerging field of network science, where one studies the structure and characterization of real-world graphs as found in social networks, WWW link structures or citation networks. A particular interesting and yet apparently under-explored direction is found in the integration of the *arrow of time* in a graph characterization, where the construction of a graph is described as an empirical process. For example in the case of citation graphs, the view that scientific research occurs in cliques in time and space, would give rise to a constantly shifting clustering of literature, and would provide as such a way to detect emerging trends in research.

Interest in this novel area of learning was established through interaction with various people, locally (as e.g. with T. De Bie and with the bio-informatics group in our department), as well as international

(e.g. last years visit at FhG, Darmstadt with U. von Luxburg and T. Hofmann, and MPI, Tübingen, Germany, and discussions at UCL-CS, London, and at ISIS, Southampton, UK) and through interactions at various recent workshops and conferences (as e.g. MLG, Berlin and ALT, Barcelona).

IV. Applications

My past and recent research are reinforced by the driving force of questions emerging in a specific *application area*. This was in particular true for the implementation and support of the toolbox LS-SVMLab. Since the online publication in 2002, it has been used successfully for research and prototyping in specific problems in bio-informatics (M@c Beth), medical signal processing (Glycensit and detection of ovarian cancer), car manufacturing (LMS and BMW), nonlinear control applications (IPCOS and BASF) and the modeling of electricity consumption (ELIA), in finance (Mastercard), in chemometrics (BASF) and traffic modeling (TML). Experience in research for application areas, and industrial practice was gained through a number of research-industry oriented projects. The main ones are the IWT project Soft4s (2000-2002), on the formulation, analysis and implementation of a technique for simulating a complex filter in the context of chemical control (partners IPCOS, BASF). A second project concerned electricity load forecasting (together with ELIA, the Flemish main electricity distributor, 2001-2003). Originating from an engineering group, I acknowledge the vital importance of successful grant and project proposal, for reasons of financial flexibility as well as for maintaining contact with reality.

V. Teaching, Supervising and Networking

Scientific relevance of a new result becomes only apparent through the (log of the number of) people it convinces. Realizing this reinforces my concern with educational, supporting and networking issues. *Teaching* and supervising students is for me about broadcasting enthusiasm and conviction that beyond the formal language of our research lies a wide field receptive for creative thinking, much like in any traditional art. This being said, I want to stress my firm belief that the teaching of a formal framework as in learning theory is indispensable for passing on insight in the field. This includes discussion of the various mechanisms of randomness, and the derivation of standard results in the field, theoretical as well as methodological. I look in particular forward to unfold a curriculum supporting my research, the main outlooks include:

- Structuring the very recent scientific history of the field
- The discussion of a formal framework for learning (universe, random mechanism, assumptions)
- Revealing the key ideas making a machine learn well
- Elaborating on different key algorithms and implementations
- Exposing the numerous links with related fields
- Presenting the impact of key application areas and case-studies
- Exposing emerging trends.

A most convincing arguments for taking further steps in an academic environment is to be able to define, elaborate, assist in, and follow a *Ph.D. research program* together with the student. My adopted view is to guide a student through in depth study of a particular case-study towards a broad know-how of the learning toolbox. This was already informal the case for the Ph.D. research of P. Karsmaker (about speech recognition), and assisting in the research of N. Pochet (kernel methods in bioinformatics) T. Van Herpe (medical signal processing for control of glycemia levels for ICU patients), S. Van Vooren (bio-informatics and data mining), V. Van Belle (nonlinear survival analysis in medical analysis); and in the supervision of various master students (including topics on microarrays, traffic analysis, nonlinear systems, cellular automata, and others). I consider in particular the development of a user-friendly and attractive website as a major tool for broadcasting new results - making up a convenient outlet for my computer science background.

Finally, I consider successful research and scientific progress as an immediate consequence of attending key conferences, and periodically visiting related research groups.

VI. Finally...

I would describe my main personal highlights in this context as being creative and being passionate on research. My personal situation is highly flexible for traveling.

My publications indicate a shift in focus from the area of artificial neural networks and purely methodological research - a.o. in a context of system identification - towards conceptual innovative and more machine learning oriented publications. A consequence is the broad working area, indicated by the wide range of publications, projects and collaborations in the interdisciplinary research group.

The above elaboration proofs commitment to academic research, and motivation for teaching and composing a curriculum. The ability for professional communication is plainly demonstrated at regular project meetings, while attention for equal opportunities is shown during my student's years in the support of disabled people (see CV).