BOOK REVIEW
"Truth from Trash. How Learning Makes Sense"
by *Chris Thornton*
The MIT Press (A Bradford Book), April 2000
ISBN 0262201275

Reviewed by José Hernández-Orallo
Departament de Sistemes Informàtics i Computació
Universitat Politècnica de València,
Camí de Vera s/n, 46022 València (Spain).
jorallo@dsic.upv.es

This book is a risky bet. It is difficult to try to make sense from the so complex field that machine learning (ML) is today. It would be even harder if this were to be done in an informal way, avoiding mathematics as far as possible, and in scarcely more than 200 pages. And finally, to make the book amusing and enjoyable seems to me a bet on a losing horse. Even worse when the horse is made from strewn and incomplete pieces: prediction, supervised learning, nearest neighbors, Kepler's Laws, redundancy, clustering, decision-tree learning, encryption, Bletchley Park, similarity, relational learning, philosophy of induction, compression, Occam's Razor and creativity. But things are not so predictable as they look at first glance, and a good rider can come up to us as a bolt from the blue. Chris Thornton has been able to combine these a priori eclectic elements in a coherent, enlightening and easy-to-read canter.

The journey runs from a short reference to the disappointment of robotics, which have been unable to endow robots with the ability to learn, to the firm rebuttal of mystical and pseudo-scientific arguments against the possibility of intelligence and creativity in machines. As the title pun suggests, the book focuses on clarifying the reasons and implications of two issues which are not well realized by ML beginners: how difficult learning is, and how ubiquitous it is for constructing our view of the world. And the tool for this clarification is the illustration of the most successful techniques and paradigms from ML. They are usually being preceded by either historical or philosophical background and followed by their applicability and limitations. In this way, the author intertwines some literary passages in form of cunning dialogues, historical anecdotes and even sly games with the reader, and lightened but still substantial technical material.

Chapter 1 is apparently bait to draw in the reader. It introduces an imaginary conference where professor A presents “the machine that can learn anything”. The context of the story is used to let some basic ideas appear, such as supervised learning, concept learning, classification and behavior learning. The author’s thesis (and professor A’s) is that these cases can be understood as special cases of prediction (of a class, which can be a Boolean value or the actuators of a robot). However, the ability to learn everything is also based on the assumption of success on “the majority of cases” but this majority is understood with just anything more than 50% of the cases. This uncovers professor A’s bluff and highlights the need for (better) selection criteria in order to completely specify the problem of induction. Nonetheless, the author’s thesis “learning = prediction” is maintained all throughout the book, without even a brief discussion on how unsupervised problems or descriptive learning problems can be transformed to supervised ones, and how this transformation may affect the difficulty (and naturalness) of the problem.

Chapter 2 begins with some word games about how easily humans find regularities from data, even when the data are noise or nonsense. I could not have imagined a better example of this than Nostradamus ‘prophecies’. However, a historical inaccuracy can be found in one of them ‘suggesting’ something about Spanish Civil War. Thornton affirms that “Francisco Franco and José Primo de Rivera were the two main leaders during Spanish Civil War” and there was “some sort of conflict between” them. Quite to the contrary, Rivera’s ‘Falange’ movement did support the rebellion of Franco in 1936 and after Rivera’s death, during the war, he was honored as a national hero by Franco. Nonetheless, the ‘prophecies’ are splendidly employed to realize the human bias towards overfitting as well as an incentive to present some well-founded learning techniques, such as nearest-neighbor classification methods.

Chapter 3 presents Science as a paradigmatic kind of induction and, hence, of learning. Thornton centers on Kepler’s astronomical theories. But, unlike many books on philosophy of science, both successes and *failures* are discussed. For instance, Kepler’s wrong but harmonious theory of geometrical concentric spheres is originated from a casual similarity between the number of perfect geometric forms (five) and the number of planets known in Kepler’s times (six). On the other hand, the discovering of the more accurate Kepler’s law of the motion of planets is also considered, as expected. A good question raised by Thornton is who is to blame in the first case and who is to congratulate in the second, the data or the analyzer of the data. In the first case a wrong number of planets and in the second case Tycho Brahe’s data set of astronomical data, extremely precise for the time.

Chapter 4 is a starter on information theory, including definitions of uncertainty, negentropy, redundancy and regularity, mostly based on Shannon’s communication theory and symbol probabilities. However, Thornton ends up the chapter equating “learning with redundancy elimination, which, in technical terms, is the task of data compression”, although the implication of this connection is postponed until chapter 10.

Chapter 5 is relatively more technical and includes a summary of fence-and-fill algorithms, such as k-means clustering, perceptron learning, multilayer ANN backpropagation, radial-basis functions, Quinlan's ID3 and C4.5, the Naive Bayes classifier and a hybrid between LVQ and C4.5 called Center Splitting. Instead of making an exhaustive and detailed account of each of them and their multiple variants (which would be almost impossible nowadays), the relevance is judiciously put on emphasizing the similarities and connections between these methods as well as their limitations.

Chapter 6 is a delightful historical interlude about the allied decryption efforts against German Enigma secret communications during Second World War, carried out by the Turing's team at Bletchley Park. Although the story is very well-known, it is used to remind us about some classical and popular concepts concerning cryptography, and the interesting and important connection between cryptography and learning, although only outlined at a shallow level. In the same vein, chapter 7 commences with an imaginary dialogue between Turing and Kepler about their particular problems. It uses the analogy between decryption and learning to drop some hints about the complexity of learning.

The rest of chapter 7 and the bulk of chapter 8 introduce one of the main theses of this book. Thornton tries to establish the point of real difficulty of learning somewhere in between non-relational learning and relational learning, the first characterized by a generally finite search space whereas relational learning has an infinite search space. The methods described in chapter 5 are then ascribed to non-relational learning. This difference can be simplistic to ILP practitioners [6], due to the different difficulty of subsets of relational learning studied in ILP (*ij*-determinations, flattened logic programs, non-recursive vs. recursive, with or without invented predicates, etc.). Nonetheless, Thornton has been successful in illustrating what the results are when addressing a relational problem with non-relational tools. Some non-relational issues of the problem can be extracted but the rest remains as noise, hence generating a poor predictive model. The nonalignment of robot sensors is a good example of the need for relational learning. He introduces the so-called "geometric separability index" to recommend which particular learning strategy should be used on raw data. Although one should refer to the literature to see how well it works in practice, the importance put on this kind of heuristic is extremely instructive. The author also reviews some of the techniques that are frequently used to apply some non-relational tools to relational problems, such as supercharging. However, they are unlikely to achieve good generalizations, and some purely relational methods such as pick-and-mix learning (used by the Bacon system) and the ILP system FOIL are also discussed as better solutions.

In Chapter 9, Thornton tells the story of a broker who is caught by the stock markets' crash of '87, partially due to the use of a ML program for making investments. This plot and some classical philosophers of induction such as Bacon, Hume, Mill and Popper, are employed to suggest that hypotheses can be falsified but never completely confirmed, placing the reader at the apex of the 'scandal' of induction. The author's goal is clear: to definitely drive the discussion towards the more comfortable view of learning as compression.

Consequently, chapter 10 introduces Kolmogorov complexity as a way to know whether a hypothesis is valuable, in a quite informal but still precise way. Then Thornton is able to shape the view of learning as removing redundancy or, in other words, compressing information. The notion of randomness and the MDL (Minimum Description Length) principle are introduced under Kolmogorov complexity. In a reverse of the usual order, the MDL principle is justified by a particular reformulation of Occam's Razor (shorter hypotheses are more plausible because they assume less). A reference (especially the extremely good one [4]) would be appreciated by avid readers after this chapter.

In Chapter 11 Thornton presents his approach for a progressive learning from non-relational to relational. His particular approach has a much too general name, "Truth From Trash" (TFT), and it is based on the "recursive" (I would say iterative) application of a fence-and-fill method followed by a recoding operation that should increase global organization. It is a nice proposal and it is apparently better than supercharged fence-and-fill. But I find some problems with its presentation in this book. First, it is not so novel and different from other proposals (incremental learning, re-description, self-improvement, etc.) to be distinguished as the "truth from trash" method. Secondly, it does not show that his proposal is better than others and that it is less computationally expensive than ILP approaches. Only a tiny and unfinished example is shown for one of the TFT instances, the so-called split-centers-in-layers (SCIL) method. The chapter concludes with a defense of the "nouvelle AI" but without disregarding classical representational approaches, somehow placing his TFT in between.

Chapter 12 intends to deal with creativity as a higher level emergent property of learning, comparing it with other emergent properties such as consciousness, awareness and aesthetics. In my opinion, creativity is much more related to explicitness/implicitness of the hypothesis with respect to the data, and the expectations of the learner and its context. In fact, some simple single-level learning systems have come out with original and creative hypotheses. I guess the author's aim was to make the reader realize (if s/he was not aware of it) that there is no magic in learning or creativity. Once mysticism is wiped out, the book is finished in a quite confident way: learning is just a hard, progressive and heterogeneous computational problem that still requires a lot of work to be done in many fields.

A vital question has not been posed in the book: now that we have learned a good hypothesis from trash, what are we going to do with it? No comment on belief or theory revision is made throughout the book, although certainly there can be no truth construction without it. Obviously, as a non-technical and short book with many literary interludes, there are many other important topics that are missing. But most of them cannot be criticized because the author's intention and the reader expectation are not those of a textbook. However, a non-technical book can be much more useful if some good references are given for further reading (especially some foundational references such as Gold's identification in the limit [2], Solomonoff's inductive inference theory, Angluin's query-based learning [1], Valiant's PAC model [9], etc.).

Having said this, the book is especially useful for pre-graduate students, for researchers of pure AI, for philosophers of science or for cognitive scientists, and, in general, for anyone who wants to enter the field of ML. Probably it cannot be used as a textbook for multidisciplinary graduate courses on ML, but it can be strongly recommended as a secondary reading. Equally, it can shrewdly be chosen as a prior and motivational reading for paving the way to more technical and comprehensive literature, such as Mitchell's [5], Langley's [3] or even Thornton's previous works [7][8]. And it can surely be used too for experienced researchers in ML to refresh some of their ideas and take a pleasing and broad-minded interlude between technicalities.

References

[1] Angluin, D. Queries and concept learning. *Machine Learning*, 2, No. 4, 319-342, 1988.

[2] Gold, E.M. Language Identification in the Limit. *Information and Control*, 10, pp. 447-474, 1967.

[3] Langley, P. *Elements of Machine Learning.* Morgan Kaufmann Publishers Inc, San Francisco, 1996.

[4] Li, M.; Vitányi, P. *An Introduction to Kolmogorov Complexity and its Applications*, 2nd Ed. Springer-Verlag 1997.

[5] Mitchell, T.M. *Machine Learning.* McGraw-Hill International Editions, 1997.

[6] Muggleton, S.; De Raedt L. Inductive Logic Programming - theory and methods. *Journal of Logic Programming*, 19-20:629-679, 1994.

[7] Thornton, C. *Techniques in Computational Learning: An Introduction.* Chapman and Hall, London, 1992.

[8] Thornton, C.; Du Boulay, B. *Artificial Intelligence: Strategies, Applications and Models Through Search.* Glenlake, 1998.

[9] Valiant, L. A theory of the learnable. *Communications of the ACM*, 27(11), 1134-1142, 1984.