Phase Transitions in Machine Learning

Phase transitions typically occur in combinatorial computational problems and have important consequences, especially with the current spread of statistical relational learning and of sequence learning methodologies. In *Phase Transitions in Machine Learning* the authors begin by describing in detail this phenomenon and the extensive experimental investigation that supports its presence. They then turn their attention to the possible implications and explore appropriate methods for tackling them.

Weaving together fundamental aspects of computer science, statistical physics, and machine learning, the book provides sufficient mathematics and physics background to make the subject intelligible to researchers in the artificial intelligence and other computer science communities. Open research issues, suggesting promising directions for future research, are also discussed.

LORENZA SAITTA is Full Professor of Computer Science at the University of Piemonte Orientale, Italy.

ATTILIO GIORDANA is Full Professor of Computer Science at the University of Piemonte Orientale, Italy.

ANTOINE CORNUÉJOLS is Full Professor of Computer Science at the AgroParisTech Engineering School, Paris.

Phase Transitions in Machine Learning

LORENZA SAITTA
University of Piemonte Orientale, Italy

ATTILIO GIORDANA
University of Piemonte Orientale, Italy

ANTOINE CORNUÉJOLS
AgroParisTech Engineering School, Paris, France

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Tokyo, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521763912

First published 2011

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-76391-2 Hardback

Cambridge University Press
978-0-521-76391-2 - Phase Transitions in Machine Learning
Lorenza Saitta, Attilio Giordana and Antoine Cornuéjols
Frontmatter
More information

Contents

Preface *page* ix
Acknowledgments xiii
Notation xiv

1 Introduction **1**

2 Statistical physics and phase transitions **12**
2.1 Basic notions of statistical physics 12
2.2 Ensemble of states 19
2.3 Phase transitions 23
2.4 Ising models 26
2.5 Mean field theory 32
2.6 Quenched disorder and self-averaging 33
2.7 Replica method 37
2.8 Cavity method 39
2.9 Comments 42

3 The satisfiability problem **43**
3.1 General framework 43
3.2 Random graphs 45
3.3 The SAT problem 49
3.4 The random $(2 + p)$-SAT 62
3.5 Solving the SAT problem 63
3.6 Comments 68

4 Constraint satisfaction problems **70**
4.1 Algorithms for solving CSPs 73
4.2 Generative models for CSPs 79
4.3 Phase transition in a CSP 81
4.4 Comments 89

5 Machine learning **92**
5.1 Concept learning 93
5.2 Representation languages 106
5.3 Comments 122

6 Searching the hypothesis space **124**
6.1 Guiding the search in the hypothesis space 125
6.2 FOIL: information gain 131
6.3 SMART+: beam search 132
6.4 G-Net: genetic evolution 133
6.5 PROGOL: exhaustive search 134
6.6 Plateaus 135
6.7 Comments 139

7 Statistical physics and machine learning **140**
7.1 Artificial neural networks 140
7.2 Propositional learning approaches 152
7.3 Relational learning 163
7.4 Comments 167

8 Learning, SAT, and CSP **168**
8.1 Reducing propositional learning to SAT 168
8.2 Phase transitions and local search in propositional learning 175
8.3 The FOL covering test as a CSP 178
8.4 Relation between CSP and SAT 179
8.5 Comments 183

9 Phase transition in FOL covering test **184**
9.1 Model RL 185
9.2 The search algorithm 195
9.3 Experimental analysis 198
9.4 Comparing model RL with other models for CSP generation 202
9.5 Smart algorithms for the covering test 214
9.6 Comments 217

10 Phase transitions and relational learning **220**
10.1 The experimental setting 221
10.2 Experimental results 229
10.3 Result interpretation 235
10.4 Beyond general-to-specific learning strategies 247
10.5 Comments 255

11 Phase transitions in grammatical inference **258**
- 11.1 Learning grammars 258
- 11.2 Grammatical inference by generalization 269
- 11.3 A phase transition in learning automata? 274
- 11.4 The covering test: random sampling in $\mathcal{H}$ 275
- 11.5 Learning, hypothesis sampling, and phase transitions 278
- 11.6 Consequences of the behavior of the learning algorithms: how bad is it? 293
- 11.7 Comments 298

12 Phase transitions in complex systems **300**
- 12.1 Complex systems 301
- 12.2 Statistical physics and the social sciences 304
- 12.3 Communication and computation networks 309
- 12.4 Biological networks 310
- 12.5 Comments 311

13 Phase transitions in natural systems **313**
- 13.1 Comments 317

14 Discussion and open issues **319**
- 14.1 Phase transitions or threshold phenomena? 320
- 14.2 Do phase transitions occur in practice? 327
- 14.3 Blind spot 329
- 14.4 Number of examples 331
- 14.5 Machine learning and SAT or CSP solvers 331
- 14.6 Relational learning and complex networks 333
- 14.7 Relational machine learning perspective 334

Appendix A *Phase transitions detected in two real cases* 339
- *A.1 Mutagenesis dataset* 339
- *A.2 Mechanical troubleshooting datasets* 347

Appendix B *An intriguing idea* 351

References 355

Index 375

Preface

From its inception in the 1930s, the rich and vigorous field of computer science has been concerned with the resources, both in time and in memory, needed to carry out a computation. A number of fundamental theorems were discovered that resorted to a worst-case analysis. The central question was whether a given algorithm could be guaranteed to terminate a computation in finite time whatever the inputs, and, if so, in which class of complexity it lay, given the control parameters: polynomial, exponential, and so on. Therefore, in 1991, a paper by Cheeseman, Kaneefsky, and Taylor came as a bolt from the blue. Indeed, while its title, "Where the really hard problems are", was not altogether disturbing, its content was. Broadly speaking, the authors argued that even if it was important to analyze worst cases, it was just as essential to look for the typical complexity of computations, the complexity encountered when solving typical problems. And there lies a gem: the transition from the region of problems that are hard, in terms of algorithmic complexity, to the region of problems that are easy can be quite sharp. Moreover, these regions and transitions are not related to the worst cases.

We remember that this 1991 paper, presented at the International Joint Conference on Artificial Intelligence (IJCAI), started a commotion, though how profound this would be was not at first apparent. We were among those who felt that this paper and others that promptly followed, from physicists in particular, were significant beyond the obvious. However, this event did not alter the course of machine learning, our field, for many years. In machine learning too the theoretical analysis that was at that time taking shape dealt with a type of worst-case study; this new statistical theory of learning was sweeping the field and gaining momentum as new learning algorithms, inspired in part by its lessons, were devised.

Thus, it was only in 1999 that M. Botta and two of us[1] finally published a paper that took in the new perspective opened by Cheeseman and others and

[1] Attilio Giordana and Lorenza Saitta (Botta *et al.*, 1999).

examined its impact on machine learning or, to be more specific, on the matching problem that is at the heart of learning. And here again, as with hindsight could have been suspected, a phase transition came into view. Over the following years we, and others, carried out thorough empirical investigations, which all confirmed and elaborated this finding. Even though the mainstream of machine learning was still under the spell of the statistical and worst-case-analysis perspective, it was becoming apparent that these results, which could not be accounted for by the dominant view, had a quite significant potential impact on the very feasibility of learning. Indeed, some known failures in the learning of large-scale problems could be explained thanks to this new point of view.

The fact is that, at least for some learning problems, there exists a sharp discontinuity between easy and hard matching problems. This severely hinders, at the very least, the assessment of candidate hypotheses considered during learning, therefore making the exploration of solutions all but blind. It is no wonder that the consequences can be quite serious.

While a complete understanding of the phase transition in learning still eludes us as a community of researchers, we feel that the wealth of results obtained in recent years and their known links with other fields in computer science and physics are now sufficiently mature to deserve a wide encompassing presentation, one that would describe as large a part as possible of the phase transition phenomena relevant to machine learning and would stimulate further research on this important subject. This book is the result of our conviction that the study of phase transitions in machine learning is important for the future of machine learning, and it presents us with the opportunity to establish profound connections with other natural sciences.

The book deals with the border between statistical physics, complex systems, and machine learning: it explores emergent properties in relational machine learning using techniques derived from statistical physics. More generally, the book is concerned with the emergence, in learning, of a phase transition, a phenomenon typically occurring both in many-body systems and in combinatorial problems.

This phenomenon is described in detail, and the extensive experimental investigation that supports its presence is reported. Then the results and the implications that the appearance of a phase transition may have on the scalability of relational learning and on the quality of the acquired knowledge are discussed in depth. With the current spread of statistical relational learning methodologies this topic is assuming an increasingly strong relevance.

The idea behind the book is to stimulate synergic research interests in the fields of both statistical physics and machine learning. Researchers in the former may find in machine learning a rich, appealing field, where their theories and methods can be applied whereas researchers in the latter may find new tools for investigating and explaining learning processes in depth.

The identification of a phase transition in a computational problem may have important consequences in practice. In fact, as mentioned above, the standard notion of the computational complexity of a class of problems is a pessimistic evaluation based on a worst-case analysis. The investigation of phase transitions can provide information on single instances of the class, shifting the focus of the complexity analysis from the maximum complexity to a *typical* complexity. Relational learning is a task particularly affected by the problem of high computational complexity. In this book, we are only concerned with supervised learning for classification, within the paradigm of *learning from examples*.

A theoretical approach, inspired by statistical physics, and a supporting set of experiments have uncovered that, in relational learning, the expected phase transition occurs inside a range of parameter values that is relevant for practical learning problems. It is thus sensible to investigate the phenomenon and to try to propose possible ways around it, since the emergence of a phase transition in relational learning can have a large negative impact on a task's feasibility.

In order to underline that the emergence of a phase transition is far from exceptional, we have widened the scope of the book to include grammar induction and an overview of related topics in neural networks and other propositional learning approaches showing the ubiquity of the phenomenon. Moving outside the machine learning area, we also describe the emergence of phase transitions in complex networks and in natural systems, including human cognition. Again, the links between the findings observed in such a variety of systems may stimulate cross-correlations and convergence.

We hope that the deep interactions that we will discuss between the theoretical issues and the experimental findings will provide a rather complete landscape of the field, including both the foundational aspects and the practical implications. Our intention is that the novelty of the topic, the analysis of foundational issues in machine learning, and our attention to practical solutions and applications will make the book appeal to a variety of readers. The detailed explanations of findings should facilitate understanding of the various viewpoints even for readers not within the field.

Even though the book mainly targets a readership familiar with artificial intelligence and machine learning, its foundational aspects will also be of interest to cognitive scientists, and even philosophers, looking for the emergence and the epistemological impact of similar phenomena in nature. The book may be of particular interest to researchers working on complex systems, as we make an explicit effort to link the phenomena investigated to the theory of these systems. Likewise, researchers in statistical physics who are interested in its computational aspects may be attracted by the book.

The approach taken is primarily quantitative and rigorous. Nevertheless, we have provided intuitive and qualitative illustrations and explanations of the issues and results. The idea is that even non-technical readers should be able to understand the main issues. For a qualitative understanding, the basic notions of artificial intelligence (especially knowledge representation and search) and computer science (especially computational complexity) are necessary. For a quantitative understanding, probability theory and advanced calculus are required.

Reading the book should allow a researcher to start work in the field without searching, reading, and linking many articles found dotted about in a variety of journals. Also, the book should be of help for those wanting to understand some of the philosophical problems underlying computation.

Above all else we would be happy to see new research themes originating from this book.

Acknowledgments

Several friends and colleagues have contributed to this book, directly or indirectly, and we are indebted to them all.

We want to thank in particular Michèle Sebag: long and pleasant discussions with her greatly contributed to our understanding and broadened our horizons; working with her was an enriching pleasure.

Also many thanks go to Erick Alphonse and Omar Osmani, who followed our first steps in the field with enthusiasm, contributing many new ideas and results.

In this book we have reported the work of many researchers, and their permission to reprint graphics from their papers has spared us a lot of additional work; hearty thanks to them as well.

Finally, we are very grateful to Enrico Scalas, a colleague physicist who carefully checked our introduction to statistical physics and took the time to provide us with detailed and insightful comments.

Notation

$\mathtt{P}$	Probability (for a finite set)
$\mathtt{p}$	Probability density
$\boldsymbol{G}$	Graph
$\mathcal{G}$	Ensemble of graphs
$\mathbb{E}[x]$	Expectation of x
$\mathbb{V}[x]$	Variance of x
$\mathcal{O}(\cdot)$	"Big O" notation: describes the limiting behavior of a function when the argument tends towards infinity
$\mathbb{R}$	The real numbers
$\mathbb{R}^n$	The space of real numbers of dimension n
$\mathbb{N}$	The natural numbers
$\mathbb{B}^n = \{0,1\}^n$	Boolean space of dimension n
$\vec{\mathbf{x}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$	A column vector
$\vec{\mathbf{x}}^\top = (x_1 \cdots x_n)$	A row vector
$\lvert\lvert \vec{\mathbf{x}} \rvert\rvert$	L_2 norm of the vector $\mathbf{x}$
$\partial/\partial x f(x,y)$	Partial derivative of function $f(x,y)$ with respect to x
$\dot{x}$ or $\mathrm{d}x/\mathrm{d}t$	Total time derivative of x
$\frac{\mathrm{d}f(x)}{\mathrm{d}x}$	Total derivative of function $f(x)$ with respect to x
$\mathcal{X}$	Input or description space of the examples
$\mathcal{Y}$	Output or label space
$\mathcal{S}_L$	Learning set

$\mathcal{P}$	Set of positive training examples
$\mathcal{N}$	Set of negative training examples
$\mathcal{S}_T$	Test set
e	An example
$\vec{\mathbf{x}}_k$	An example description
$\vec{\mathbf{z}}_k = (\vec{\mathbf{x}}_k, y_k)$	A labeled example (description, class)
$y_k \in \mathcal{Y}$	The true label of an example provided by an “oracle”
$\mathbb{C}$	Space of possible target concepts
$c : \mathcal{X} \rightarrow \mathcal{Y}$	A target concept
$\mathcal{H}$	Hypothesis space
$h \in \mathcal{H}$	A hypothesis considered by the learner
$y = h(\vec{\mathbf{x}}) \in \mathcal{Y}$	Prediction of the hypothesis h about example $\vec{\mathbf{x}}$
Φ	A set of logical formulas
$\varphi \in \Phi$	A logical formula φ belonging to the set Φ
$\ell\big(c(\vec{\mathbf{x}}), h(\vec{\mathbf{x}})\big)$	The loss incurred when $h(\vec{\mathbf{x}})$ is predicted instead of the true label $c(\vec{\mathbf{x}})$
E	Energy
S	Entropy
$x_j{:}a$	Variable x_j is bound to the constant a
$\overline{x_j{:}a}$	Variable x_j is not bound to the constant a
m	Number of literals in a formula to be matched
n	Number of variables in a formula to be matched
N	Number of goods in each table in an example
L	Number of constants occurring in the tables of an example