Summary of "Restricted Boltzmann Machines and the Renormalization Group: Learning Relevant Information in Statistical Physics"

Jesse Q. Hoogland
Amsterdam University College
Amsterdam, the Netherlands
jessequinten@gmail.com
Dr. P. Marcos Crichigno
Supervisor
University of Amsterdam
Amsterdam, the Netherlands
P.M.Crichigno@uva.nl

In the age of *big data*, machine learning (ML), a subset of artificial intelligence (AI), has become more than *just* another set of data analysis tools [1]. For one, ML's connections with theoretical physics are multivarious and deeply conceptual; the very success of ML may, in part, result from physical principles including symmetry, locality, and hierarchy [2]. Furthermore, ML and theoretical physics share a powerful conceptual framework in information theory [3]. Beyond data analysis, the intersection of ML and physics contains a unique set of ideas that researchers in both fields can leverage to solve tough problems.

In particular, recent work has drawn attention to the similarities between ML and a class of techniques from statistical physics known as the renormalization group (RG) [4, 5, 6, 2]. Developed in the last century, RG has been crucial in making sense of critical behavior, those phenomena characterizing phase transitions. In 2014, Mehta and Schwab published a seminal paper describing an *exact* equivalence between a technique from RG and a type of neural network (NN) from ML [4]. This, however, met criticism, and it works only under a narrow set of circumstances. The similarities between ML and RG, then, are still largely qualitative, and this remains an active area of research. Not to mention, the research landscape maintains lingering misconceptions about the details of the intersection [6, 2, 7, 8]. This warrants further investigation, and in order to facilitate and encourage such research, our first contribution is to provide a clarifying overview of the competing views. We resolve a number of inaccuracies.

In 2018, Koch-Janusz and Ringel derived an algorithm which uses neural networks to learn RG transformations on lattice systems: the *real-space mutual information* (RSMI) algorithm [5]. Notably, this method is *unsupervised* which is particularly relevant for research into poorly understood physical systems; information-theoretic approaches, like this algorithm, may guide researchers towards the locations of critical points and even calculations of critical exponents. Furthermore, Koch-Janusz and Ringel's derivation is *optimal* in a rigorous sense [5, 9]. This is exciting because many well-established practices in RG lack precise justification. More exact formulations, like these, may inspire more effective implementations, not to mention a better understanding of why these ML and RG techniques work.

In our investigation, we will develop a set of tools for tackling critical phenomena. First, we consider some of the standard techniques of statistical physics, building towards an ML-derived implementation of RG. We, then, introduce elements of ML, emphasizing their utility in a variety of statistical physics contexts. We anchor this investigation around the Ising model, one of the most important models in statistical physics. To compare these various techniques, we evaluate their ability in predicting the Ising model's correlation length critical exponent, ν.

To accomplish this, we have built and shared an open-source implementation of the RSMI algorithm [10] through *rgpy*. Hereby, we provide a calculation for ν.Then, we describe a generalization of this algorithm to *arbitrary* lattice systems. This gives rise to a family of RSMI-inspired approaches.

It is our aim to enable and inspire researchers to build further on our results. We accomplish this by reviewing the current state of research, sharing an implementation of the RSMI algorithm, and describing avenues of future research. Although we focus on the perspective of statistical physicists, this capstone is accessible for both ML and physics researchers, even at an undergraduate level.

In chapter 2, we begin by introducing techniques native to statistical physics. We describe mean-field theory, and its failures bring us to the renormalization group. In chapter 3, we discuss two examples of ML in physical investigations. First, we use neural networks to classify Ising model phases. Then, we use the same neural networks to generate new samples of Ising models. These examples serve to introduce the basics of ML, assuming no prior knowledge (except mathematical maturity), and the same is true for the portion on statistical physics. In chapter 4, we explore the similarities between ML and RG, and by being explicit in our formulation of "relevant" information, we manage to avoid some of the mistakes of earlier comparisons. In chapter 5, we explain and justify the RSMI algorithm, following the formulation of Koch-Janusz and Ringel. In chapter 6, we provide our own results: a recalculation of ν and a generalization of this technique, paving the way for a new class of RG techniques. In chapter 7, we close with a discussion, reflecting on our comparisons of ML and statistical physics and emphasizing the wide-ranging impacts of these ideas.

Bibliography

[1] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436, 2015.

[2] Henry W. Lin, Max Tegmark, and David Rolnick. Why does deep and cheap learning work so well? *Journal of Statistical Physics*, 168:1223–1247, Sep 2017.

[3] Naftali Tishby, Fernando C Pereira, and William Bialek. The information bottleneck method. *arXiv preprint physics/0004057*, 2000.

[4] Pankaj Mehta and David J. Schwab. An exact mapping between the variational renormalization group and deep learning. *arXiv e-prints*, page arXiv:1410.3831, Oct 2014.

[5] Maciej Koch-Janusz and Zohar Ringel. Mutual information, neural networks and the renormalization group. *Nature Physics*, 14:578–582, Jun 2018.

[6] Satoshi Iso, Shotaro Shiba, and Sumito Yokoo. Scale-invariant feature extraction of neural network and renormalization group flow. *arXiv e-prints*, 97:053304, May 2018.

[7] David J. Schwab and Pankaj Mehta. Comment on ”why does deep and cheap learning work so well?” [arxiv:1608.08225]. *arXiv e-prints*, page arXiv:1609.03541, Sep 2016.

[8] Jaco ter Hoeve. Renormalization group connected to neural networks, 2018.

[9] Patrick M. Lenggenhager, Zohar Ringel, Sebastian D. Huber, and Maciej Koch-Janusz. Optimal renormalization group transformation from information theory. *arXiv e-prints*, page arXiv:1809.09632, Sep 2018.

[10] Jesse Hoogland. rgpy. `https://github.com/jqhoogland/rgpy`, 2019.