氏　　　名	横尾　純斗
学位(専攻分野)	博士(理学)
学位記番号	総研大甲第 2151 号
学位授与の日付	2020 年 3 月 24 日
学位授与の要件	高エネルギー加速器科学研究科　素粒子原子核専攻 学位規則第6条第1項該当
学位論文題目	Feature extraction of two dimensional Ising model by unsupervised neural networks
論文審査委員	主　査　　講師　山田　憲和 教授　北澤　良久 准教授　西村　淳 講師　板倉　数記 教授　磯　暁 教授　松枝　宏明 仙台高等専門学校　情報電子システム工学専攻

（様式３）

博士論文の要旨

氏　名　横尾　純斗

論文題目　Feature extraction of two dimensional Ising model by unsupervised neural networks

Machine learning is a technique which makes computers perform various intellectual tasks such as discrimination of subtle images, translation of languages, or generation of images. This technique has been very successful and received considerable attention in data science as well as natural science. This success comes from deep learning which is one of the models to achieve machine learning with Deep Neural Network (DNN) calculation motivated by the neural network of biological brain. It was considered that the DNN has too many parameters to calculate, but the recent development of computer makes the calculation possible.

It is true that the DNN is a very useful framework, but why does it work so well? In order that machine learning works successfully, it is crucial to find suitable representations of input data for tasks we want to achieve. In general, however, it is very difficult to design the suitable representations by human hands. One of the ideas for overcoming this problem is to get the representations by machine learning. This is called representation learning. It is believed that DNN automatically achieves this representation learning in the process of its training. In other words, DNN can extract "features" from complicated dataset effectively.

However, the theoretical understanding of how DNN extracts the features from dataset is unclear. One of possible answers to these questions is following. Input data and their specific features usually have a hierarchical structure: an image of a cat can still be identified as an animal in a very low resolution image but you may not be able to distinguish it from a dog. Such hierarchy of features can be efficiently reflected by the deep structure of neural networks. Namely, it is believed that DNN learns low-level (microscopic) representations in the upper stream of the network and gradually extracts higher-level (macroscopic) representations as the input data flows downstream. In other words, the initial data will get coarse-grained towards the output. This view is reminiscent of the renormalization group (RG) in statistical physics and quantum field theories, and various thoughts and studies are given based on this analogy.

The RG is the most important concept and technology to understand the critical phenomena in statistical physics and also plays an essential role to constructively define quantum field theories on lattice. It is based on idea (and proved by Kenneth Wilson) that the long-distant macroscopic behavior of a many body system is universally described by relevant operators (relevant information) around a fixed point, and not affected by microscopic details in the continuum limit. Through reduction of degrees of freedom in RG, the relevant information is emphasized while other irrelevant information is discarded.

In this thesis, in order to explore the feature extraction of DNN, we train Restricted Boltzmann Machine (RBM) by spin configurations of 2-dimensional Ising model and construct a flow of temperature generated by the trained RBM. This flow is motivated by the renormalization group flow of a statistical model. It is thought that this flow emphasizes the "relevant" features the unsupervised networks learn, and eliminate "irrelevant" information from the dataset. In our simulations, we provide three different types of trainings. One type of RBM (we call type V) is trained by configurations at various temperatures from low to high. Other two types (type H and L) are trained by configurations only at high (and only at low) temperatures. After we fix the parameters in RBM by training, we iterate reconstruction of the spin configurations and generate the flows of the configurations. Then, we translate these flows of the configurations into the flows of the temperature. In order to measure the temperature of a distribution of configurations, we prepare another neural network trained by a supervised learning. Our results are following. In type H/L RBM, the temperature approaches higher/lower temperature than the critical temperature as expected. On the other hand, in the type V RBM, the temperature approaches the critical point as opposed to the conventional RG flow of the Ising model. This means that type V RBM extracts the critical point as features even though we do not give the information about the phase transition. We also analyzed these results by the singular value decomposition of weight matrix.

Is it universal that the flow of unsupervised network approaches to the critical point? To check this, we investigated the flow by Autoencoder (AE). As in RBM, we train AE by configurations at various temperatures from low to high. We generate the flows of configurations by AE and translate these configuration flows to the temperature flows by supervised network. We compare the fixed point temperatures of AEs trained by different learning epochs. Unlike the RBM case, the temperature of the fixed points of AE trained sufficiently is lower than the critical temperature if AE is sufficiently trained (5000 epochs). On the other hand, if the learning epochs are 1000, the fixed point temperature monotonically increase with the number of hidden units. We found that the fixed point becomes the critical point when the gapped structure in singular value spectrum of weight matrix vanishes. We also prepare the datasets in narrow range of temperatures: higher temperatures case corresponding to type H ($T = 4.6, 4.7, \cdots, 5.0$), lower temperatures case corresponding to type L ($T = 1.0, 1.1, \cdots, 1.5$) and around the critical temperature case ($T = 2.2, 2.3$). The fixed point temperatures are higher than the critical temperature, lower than the critical temperature and around the critical temperature, respectively. This result suggests that AE can learn the specific temperature of the dataset and the fixed point of the AE flow corresponds to the features which AE learns.

Form8・Separate Sheet
（様式８・別紙）

Results of the doctoral thesis screening

博士論文審査結果

Name in Full
氏　名　横尾 純斗

Title
論文題目 Feature extraction of two dimensional Ising model by unsupervised neural networks

機械学習の手法は近年様々な分野へ応用され、新しいタイプの成果を次々と挙げている。しかしながら、未だ様々な機能がうまくいく原理について深い理解が得られているとは言い難い。物理現象の解析への応用も始まっているが、今後ブラックボックスのままでは困る場合が生じることが予想される。横尾純斗氏の学位論文では、機械学習が得意とする機能の一つである「特徴抽出」に着目し、そのメカニズムを明らかにすることを目的としている。特徴抽出では、Restricted Boltzmann Machine (RBM)と呼ばれる教師なし学習を経た”machine”がよく使われる。この”machine”は入力したデータの特徴を出力するという機能を持つ。そこで、２次元イジング模型を例にとり、ある温度の正準分布に従うように生成されたスピン配位を入力として与えると何が出力されるのか、またなぜそのように変換されるのかについて調べた。まず、様々な温度で学習したRBMを用意し変換後の配位の温度の変遷をモニターしたところ、入力した配位の温度に依らず変換を繰り返すうちに徐々に臨界温度に近づき、最終的にそこに留まることを見出した。注目すべきは、RBMには相転移の存在も相転移温度も教えていない点である。このことは機械学習が自然に２次元イジング模型の特徴を習得していることを意味し、大変興味深い発見である。この理由を探るために、学習により変化するパラメータである weight から作られる行列の特異値と固有ベクトルの分布や温度依存性を詳細に解析した。その結果、大きな特異値を持つ固有ベクトルは大きなクラスター状の空間分布をしおり、特異値が小さくなるにつれクラスターのサイズが小さくなっていくことが分かった。その他に行われた解析と総合すると、様々な温度で学習した RBM の weight は様々なスケールのクラスターを情報として内包しており、このことが２次相転移点近傍で実現されるスケール不変性を「知る」ことに繋がっていることを見出した。

当研究では、横尾氏はプログラム言語をいち早く習得し、高いスキルを身につけるに至った。また、プログラム開発と解析の全てを独力で行い、得られた数値結果をよく吟味した上で解釈を与えている様子が窺えた。非常にチャレンジングな研究目的であるにもかかわらず、ユニークな着眼点の下で、次のステップに繋がる礎となる成果をあげることができた点は評価できる。

上記の結果は学術的に十分な価値が認められるものであり、博士論文の内容として必要な水準を満たしていると判断し、横尾純斗氏の博士論文審査を合格と判定した。