Transfer Machine Learning of an Anisotropic Model

D. D. Sukhoverkhova[1, 2, *] and L. N. Shchur[1, 2, **]

(Submitted by A. M. Elizarov)

[1]Landau Institute for Theoretical Physics, 142432, Chernogolovka, Moscow region, Russia
[2]National Research University Higher School of Economics, 101000, Moscow, Russia
Received October 25, 2024; revised November 20, 2024; accepted November 30, 2024

Abstract—We investigate the possibility of extracting features of second-order phase transitions using transfer machine learning. We have performed supervised machine learning for binary classification of snapshots of the spin distribution of the isotropic Ising model. The binary classification is performed in ferromagnetic and paramagnetic phases using a known critical temperature. The trained network is used to predict whether a snapshot obtained from model simulations with orthogonal anisotropy belongs to the paramagnetic phase. Using finite-dimensional prediction analysis, we estimate the critical temperature and the exponent of the correlation length. This gives us an estimate of the interval of the anisotropy parameter in which the neural network can make correct estimates.

2010 Mathematical Subject Classification: 68U20, 68W40

Keywords and phrases: supervised machine learning, lsing model, critical behavior, universality class, transfer learning

1. INTRODUCTION

Machine learning methods have become widespread in tasks that require processing big amounts of data. Such tasks include the study of statistical physics systems, which are characterized by a large volume of phase space. With data on the thermodynamic ensemble, it is possible to make estimates of the properties of systems as a whole [1]. For example, one of the first studies in this area [2] was devoted to obtaining information about the second-order phase transition in the Ising model. The authors formulated a classification problem for a neural network: in the Ising model, the phase transition occurs between ferro- and paramagnetic phases, which define two classes. By analyzing the probability distribution of each phase, we can estimate the critical transition temperature and some universal properties of the Ising model.

Models with the same (a) dimensionality, (b) symmetry and (c) degeneracy of the ground state form universality classes. In the field of phase transitions, models from the same universality class have the same properties. With the help of machine learning (neural networks) it is possible to obtain the properties of the universality class in which the model with input data [3] is located. In particular, a method is known for estimating the growth rate of the correlation length diverging at a second-order phase transition point [4]. The growth rate, or critical exponent of the correlation length, takes the same value for all models from the same universality class. The question arises as to the feasibility and accuracy of estimating this same exponent using a neural network approach in other models from the same universality class. For this purpose, it is necessary to apply the knowledge obtained during the training of the neural network to new data.

The method of using a neural network pre-trained on one task to make predictions on another task is called transfer learning [5]. Previously, we investigated the applicability of transfer learning [6] for square and triangular Ising models in the same universality class. In this paper, we continue to study transfer learning in the same universality class, but now with the example of orthogonal anisotropy in the Ising [4] model.

* E-mail: diana.suhoverhova@mail.ru
** E-mail: lev@landau.ac.ru

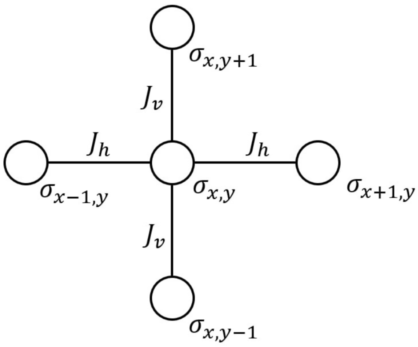

Figure 1. The spin $\sigma_{x,y}$ and its nearest neighbors in the Ising model with orthogonal couplings.

We begin by training a neural network to binary classify samples of the isotropic Ising model: the values of the coupling constants on the vertical J_v and horizontal J_h coincide (see expression (1)). The data are classified relative to the critical temperature of the isotropic model first derived in [7]. We then test the pre-trained network on anisotropic samples. The strength of anisotropy depends on the parameter $\kappa = J_v/J_h$, which we gradually decrease: $\kappa < 1$.

The output of the neural network is a prediction of whether the instantaneous snapshot belongs to the paramagnetic phase $p_i(T; L) \in [0, 1]$. We feed the neural network a set of snapshots at a given temperature, average the resulting predictions, and obtain estimates of the paramagnetic phase probability at a given temperature. We also investigate the standard deviations of the $D(T; L)$ phase probabilities.

By analyzing the functions $P(T; L)$, we obtain estimates of the critical temperature, and from the functions $D(T; L)$ we obtain the exponent of the critical correlation length. The obtained results indicate that anisotropy affects the estimates of the investigated quantities in a nontrivial way.

2. MODEL AND DATA SETS

We consider the orthogonal Ising model on the square lattice $L \times L$ (Fig. 1)

$$\mathcal{H} = -\sum_{x,y=1}^{L} \sigma_{x,y} \left[J_h \sigma_{x+1,y} + J_v \sigma_{x,y+1} \right]. \tag{1}$$

Data samples are generated by the Metropolis algorithm, each instantaneous sample is a black and white image, or snapshot. The thermalization time is $20 \times L^{2.15}$ [8]. After equilibration, each snapshot is stored once for $2 \times L^{2.15}$ Monte Carlo (MC) steps and each MC consists of $L \times L$ local flips of spins.

We fix the horizontal coupling parameter $J_h = 1$, and vary the vertical coupling parameter J_v with $\kappa = 1, 3/4, 1/2, 1/4, 1/8, 1/16$. We also consider several sizes of linear lattices $L = 20, 30, 60, 80, 100, 120$.

The value of the critical temperature as a function of J_v and J_h is known from the analytical solution of Onsager [4]:

$$\sinh \frac{2J_v}{k_B T_c} \sinh \frac{2J_h}{k_B T_c} = 1. \tag{2}$$

Each data set is generated over a range of $T_c \pm 0.3$, totaling 100 equidistant from each other temperature points in 6×10^{-3} increments. Thus, T_c varies as a function of J_v according to the formula (2), which gives the coverage of the region under study necessary to hit the phase transition temperature. At each T of the range considered, we save $N = 2048$ snapshots in the case of isotropic sampling $\kappa = 1$ for training the network. For testing anisotropic datasets $J_v \neq J_h$, we save only $N = 512$ snapshots. Thus, the total size of the dataset for each lattice size L and anisotropy parameter κ is 100 N snapshots.

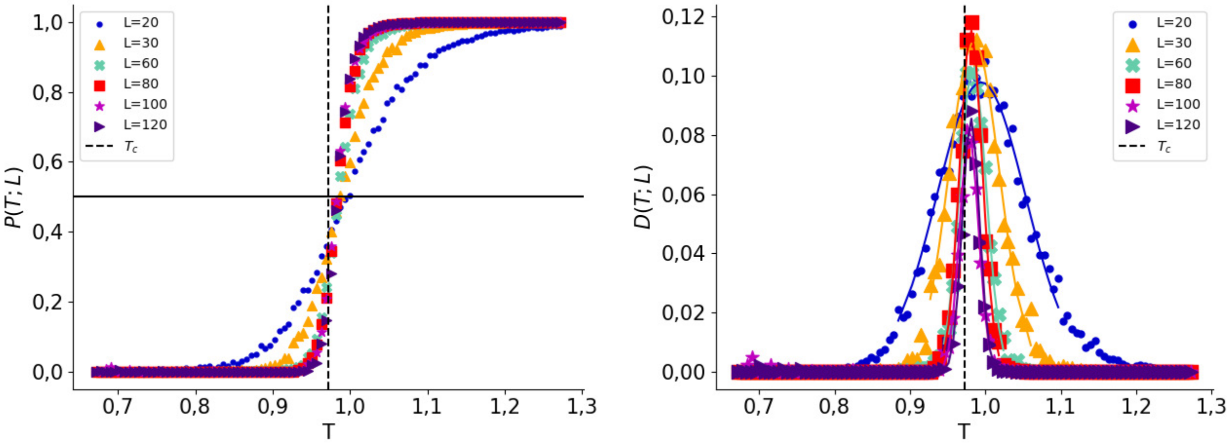

Figure 2. Left: probability function of samples belonging to paramagnetic phase $P(T;L)$ and Right: standard deviations of probability of samples belonging to paramagnetic phase $D(T;L)$. Anisotropy parameter $\kappa = 1/8$.

In addition, since the phase transition temperature T_c is known from the exact solution, we label the training data into two classes. All images acquired at temperature $T < T_c$ are labeled class 0, denoting the ferromagnetic phase, and similarly, all images acquired at $T > T_c$ are labeled class 1, denoting the paramagnetic phase [2]. Labels are assigned to training data only and are not used in testing.

3. TRAINING AND TESTING A NEURAL NETWORK

We apply a convolutional neural network (CNN) architecture consisting of one convolutional layer, two full-link layers, and ReLU activation between them [3]. We train multiple neural networks for each of linear dimension L. The networks are trained in only one epoch to avoid overtraining [9].

So, we train the neural networks on an isotropic data set $\kappa = 1$ for all values of lattice size L. Then, we test the pre-trained networks on each data set $\kappa < 1$.

When testing snapshots, the network takes as input instantaneous snapshots of size $L \times L$, and returns as output a single number $p_i(T;L) \in [0,1]$ – the prediction that the snapshot belongs to the paramagnetic phase $T > T_c$. Since the test data sets contain N samples at each temperature, we average the obtained probabilities and construct the paramagnetic phase probability function

$$P(T;L) = \frac{1}{N}\sum_{i=1}^{N} p_i(T;L).$$

Also, similar to the method [3], we calculate the standard deviations of the network outputs $p_i(T;L)$ at each temperature value T of the images under test

$$D(T;L) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i(T;L))^2 - \left(\frac{1}{N}\sum_{i=1}^{N} p_i(T;L)\right)^2}.$$

For example, the obtained functions $P(T;L)$ and $D(T;L)$ for the data set $J_v = 1/8$ are shown in Fig. 2.

4. ESTIMATION OF THE CRITICAL TEMPERATURE

Based on the functions $P(T;L)$ and $D(T;L)$, we estimate the critical temperature of the phase transition:

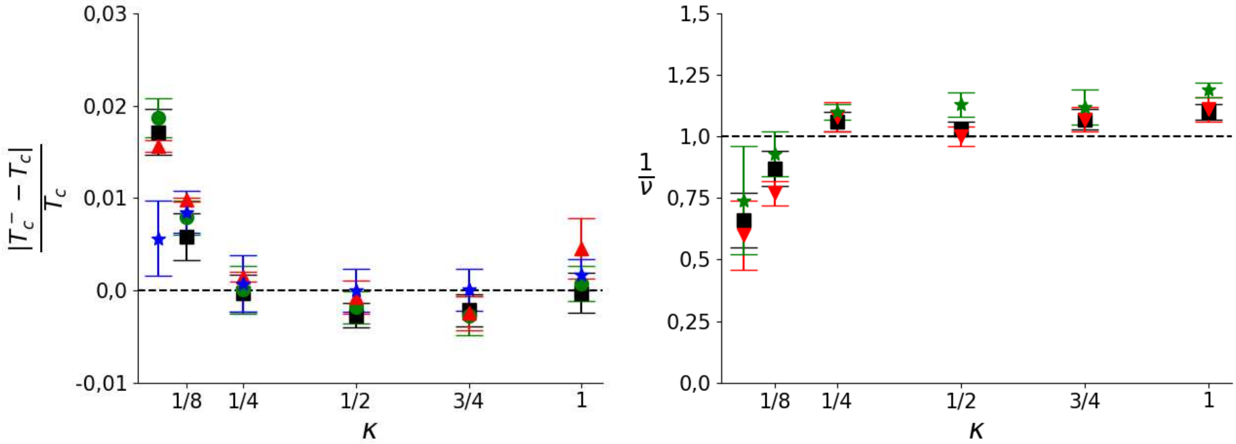

Figure 3. Left: Relative error of critical temperature estimate T_c^-: black squares are T_c^*, red triangles are T_c°, green circles are $T_c^\triangle$, blue stars are μ. Right: Estimation of the critical correlation length exponent $(1/\nu)_-$: black squares are $1/\nu$, full width; red triangles are $(1/\nu)_r$, right half-width; green stars are $(1/\nu)_l$, left half-width

1. Temperature T_c^* of a discrete set of points $T_c \pm 0.3$ at which the probability of paramagnetic phase is closest to the value 0.5
$$T_c^*(L) = \min_T |P(T;L) - 0.5| \pm 6 \cdot 10^{-3};$$
2. From the set $T_c \pm 0.3$, we select two temperature points that define the lower and upper bounds of the phase probability near 0.5; we connect the two points with a straight line. The temperature estimate T_c° is the intersection of the line with $y = 0.5$
$$T_{min} = \min_T |P(T;L) - 0.5|,\ P(T;L) < 0.5, \quad T_{max} = \min_T |P(T;L) - 0.5|,\ P(T;L) > 0.5,$$
$$P_{min} = P_{min}(L) = P(T_{min};L), \quad P_{max} = P_{max}(L) = P(T_{max};L),$$
$$T_c^\circ(L) = \frac{T_{min}(P_{max} - 0.5) + T_{max}(0.5 - P_{min})}{P_{max} - P_{min}}. \quad (3)$$
3. For several points near 0.5, the intersection of two straight lines: $y = a_1x + b_1$, fitting several (2) probability estimates $P(T;L)$ in the neighborhood of 0.5, and $y = a_2x + b_2$, fitting inverse probability estimates $1 - P(T;L)$:
$$T_c^\triangle = \frac{b_2 - b_1}{a_1 - a_2} \pm \left| \frac{a_1b_2 - a_2b_1}{a_1 - a_2} - 0.5 \right|;$$
4. Temperature based on the function $D(T;L)$: the distribution of $D(T;L)$ at each L is approximated by a Gaussian function with mean μ and standard deviation σ. We consider $\mu(L)$ as an estimate of the critical temperature.

All estimates are constructed by testing the neural network on finite lattice sizes L. We then make an estimate of the corresponding temperature in the thermodynamic limit, according to the shift [10, 11] of the critical temperature at finite dimensions $T_c^-(\infty) = T_c^-(L) + a/L$. The left panel of Fig. 3 and Table 1 show the relative errors of the obtained estimates of the critical temperature using all four methods.

5. ESTIMATION OF THE CORRELATION LENGTH EXPONENT

It is known [4] that in the universality class of the two-dimensional Ising model, the correlation length ξ diverges at the phase transition point: $\xi \propto \tau^{-1/\nu}$ with critical exponent $\nu = 1$, and $\tau = (T - T_c)/T_c$ is the reduced temperature.

κ	$-$	$T_c^-(\infty)$	T_c	$\Delta T^-/T_c$	$\Delta T^-/\sigma_T$
1	$*$	2.2685(49)	2.2692	–0.0003(22)	0.1
	$\circ$	2.2708(43)		0.0007(19)	0.4
	$\triangle$	2.2795(74)		0.0045(32)	1.2
	μ	2.2718(34)		0.0012(15)	0.8
3/4	$*$	1.9686(34)	1.9728	–0.0021(17)	1.2
	$\circ$	1.9674(42)		–0.0028(21)	1.3
	$\triangle$	1.9680(36)		–0.0025(19)	1.3
	μ	1.9709(41)		–0.0010(21)	0.5
1/2	$*$	1.6366(21)	1.6410	–0.0027(13)	2.1
	$\circ$	1.6380(28)		–0.0018(17)	1.1
	$\triangle$	1.6398(29)		–0.0007(18)	0.4
	μ	1.6388(34)		–0.0014(21)	0.6
1/4	$*$	1.2387(25)	1.2391	–0.0003(20)	0.2
	$\circ$	1.2392(32)		0.0001(26)	0.03
	$\triangle$	1.2409(7)		–0.0015(5)	2.6
	μ	1.2384(32)		–0.0005(26)	0.2
1/8	$*$	0.9780(25)	0.9723	0.0058(26)	2.3
	$\circ$	0.9800(18)		0.0079(18)	4.3
	$\triangle$	0.9819(2)		0.0098(2)	48.0
	μ	0.9799(20)		0.0078(21)	3.8
1/16	$*$	0.8019(20)	0.7884	0.0172(25)	6.8
	$\circ$	0.8031(17)		0.0187(21)	8.6
	$\triangle$	0.8007(5)		0.0156(6)	24.6
	μ	0.7921(29)		0.0047(37)	1.3

Table 1. Estimates of the critical temperature T_c^-; ΔT_c^- denotes the absolute error. The symbols are introduced in Section 4 and correspond to the temperature estimation method.

We find an estimate of the exponent ν using the hypothesis [3] that the width of $\sigma(L)$ at finite lattice sizes behaves in the same way as the width of thermodynamic functions [10, 11]: $\sigma(L) \propto bL^{-1/\nu}$. The evaluation results are shown in the right panel of Fig. 3 and in Table 2.

6. DISCUSSION

Figure 3 show that when the anisotropy parameter $\kappa = J_v/J_h$ is reduced from the isotropic value of 1 to within 1/4, the estimates of the critical temperature and critical exponent are within error with the theoretically known values of the temperature calculated from the expression (2) and the critical exponent of the correlation length $\nu = 1$. In other words, in this range of anisotropy, the neural network is able to correctly estimate the phase transition characteristics under cross-learning. In other words, the neural network correctly produces predictions.

However, as the anisotropy parameter $\kappa = J_v/J_h$ decreases, a systematically increasing deviation of both the phase transition temperature estimate and the critical exponent estimate is observed at values of 1/8 and 1/16. A possible explanation for this deviation could be purely geometric, due to a change in the lattice aspect ratio. We tested this hypothesis by varying the vertical length of the lattice in proportion to the anisotropy parameter. This led to significant changes in the estimates

κ	$\frac{1}{\nu}$	$(\frac{1}{\nu})_r$	$(\frac{1}{\nu})_l$
1	1.10(3)	1.11(5)	1.19(3)
3/4	1.07(4)	1.07(5)	1.12(7)
1/2	1.03(3)	1.00(4)	1.13(5)
1/4	1.06(4)	1.08(6)	1.10(3)
1/8	0.87(7)	0.77(5)	0.93(9)
1/16	0.66(11)	0.60(14)	0.74(22)

Table 2. Estimation of the critical correlation length exponent $(1/\nu)_{-}$: $1/\nu$ is the full width; $(1/\nu)_r$ is the right half-width; $(1/\nu)_l$ is the left half-width.

of the critical temperature and the critical exponent ν. Thus, the attribution of the deviations to the aspect ratio is untenable.

Another possible explanation can be proposed on the basis of the analysis of the correlation length behavior obtained in [12]. In this work, an expression for the correlation function of spins at radially measured distance R was obtained, which has the form:

$$\langle \sigma_{0,0}\sigma_{x,y} \rangle = F(t)/R^{1/4} + F_1(t)/R^{5/4} + o(R^{-5/4}) \tag{4}$$

and the main amplitude $F(t)$ and the correction amplitude $F_1(t)$ are functions of the recalled distance t

$$t = |z_1 z_2 + z_1 + z_2 - 1|\, R\big[z_1 z_2(1 - z_1^2)(1 - z_2^2\big]^{-1/4}$$

expressed through the variables $z_1 = \tanh(J_h/T)$ and $z_2 = \tanh(J_v/T)$.

The radial measure R is a complex distance function on the lattice including the variables z_1 and z_2, and its expression is given in formula (2.6) of the article [12]. To give some idea of its form, in the isotropic case on a lattice with horizontal length L_h and vertical length L_v it has the simple form $R = \sqrt{L_h^2 + L_v^2}$. It is important for us that the ratio of the amplitudes $F(t)$ and $F_1(t)$ changes sharply with decreasing anisotropy parameter and becomes especially noticeable when κ is less than 1/4. This is qualitatively similar to the sharp increase in the deviations of the temperature and critical index estimates in our case. The deviations in the Fig. 3 of our paper are similar to Fig. 1 of the paper [12], which shows the ratio $F_1(t)/(tF(t))$ as a function of the anisotropy parameter κ. This means that for small values of κ, the influence of $F_1(t)/R^{5/4}$ in 4 grows and the spatial behavior of the correlation function changes. As a result, the neural network makes a mistake in estimating the spatial correlations and incorrectly predicts that the images belong to the paramagnetic phase.

7. ACKNOWLEDGEMENT

Modeling and machine learning were carried out using high-performance computing facilities of the National Research University Higher School of Economics.

8. FUNDING

The study was supported by the Russian Science Foundation grant 22-11-00259.

9. CONFLICT OF INTEREST

The authors of this work declare that they have no conflicts of interest.

REFERENCES

1. G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborov, “Machine learning and the physical sciences”, Rev. Mod. Phys. 91, 045002 (2019).
2. J. Carrasquilla and R. G. Melko, “Machine learning phases of matter”, Nat. Phys. 13, 431 (2017).
3. V. Chertenkov, E. Burovski, and L. Shchur, “Finite-size analysis in neural network classification of critical phenomena”, Phys. Rev. E 108, L031102 (2023).
4. L. Onsager, ”Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition”, Phys. Rev. 65, 117 (1944).
5. K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning”, J. Big data 3, 1 (2016).
6. D. D. Sukhoverkhova and L. N. Shchur, “Influence of anisotropy on the study of critical behavior of spin models by machine learning methods”, JETP Letters 120, 619 (2024).
7. H. A. Kramers and G. H. Wannier, “Statistics of the Two-Dimensional Ferromagnet. Part I”, Phys. Rev. 60, 252 (1941).
8. V. Chertenkov and L. Shchur, “Universality classes and machine learning”, J. Phys.: Conf. Ser. 1740, 012003 (2021).
9. V. I. Chertenkov, PhD thesis, “Universality classes and machine learning”, National Research University Higher School of Economics (2024).
10. M. E Fisher and A. E. Ferdinand, “Interfacial, Boundary, and Size Effects at Critical Points”, Phys. Rev. Lett. 19, 169 (1967).
11. A. E. Ferdinand and M. E. Fisher, “Bounded and Inhomogeneous Ising Models. I. Specific-Heat Anomaly of a Finite Lattice”, Phys. Rev. B 185, 832 (1969).
12. T. T. Wu, D. M. McCoy, C. F. Tracy, and E. Barouch, “Spin-spin correlation functions for the two-dimensional Ising model: Exact theory in the scaling region”, Phys. Rev. B 13, 316 (1976).