

Влияние анизотропии на исследование критического поведения спиновых моделей методами машинного обучения

Д. Д. Суховерхова, Л. Н. Щур¹⁾

Институт теоретической физики им. Л. Д. Ландау РАН, 142432 Черноголовка, Россия

Национальный исследовательский университет Высшая школа экономики, 101000 Москва, Россия

Поступила в редакцию 16 сентября 2024 г.

После переработки 16 сентября 2024 г.

Принята к публикации 18 сентября 2024 г.

В статье мы применили глубокую нейронную сеть для изучения вопроса переносимости знания между моделями статистической механики. Был проведен следующий компьютерный эксперимент. Сверточная нейронная сеть была обучена для решения задачи бинарной классификации моментальных снимков расположения спинов модели Изинга на двумерной решетке. При тестировании на вход нейронной сети подавались моментальные снимки расположения спинов модели Изинга на решетке с диагональными ферромагнитными и антиферромагнитными связями. По выходам тестируемой сети были получены оценки вероятности принадлежности образцов к парамагнитной фазе. Анализ этих вероятностей позволил оценить критическую температуру и критический показатель корреляционной длины. Выяснилось, что при слабой анизотропии нейронная сеть удовлетворительно предсказывает точку перехода и значение показателя корреляционной длины. Сильная анизотропия приводит к заметному отклонению предсказанных величин от точно известных. Качественно сильная анизотропия связана с наличием осцилляций корреляционной функции выше температуры возникновения беспорядка по Стефенсону и дальнейшим приближением к точке полностью фрустрированного случая.

DOI: 10.31857/S0370274X24100239, EDN: VVMLUH

1. Введение. Машинное обучение начинает формировать в последнее десятилетие дополнительный подход в физических науках в дополнение к традиционным подходам – эксперименту, теории и компьютерному моделированию [1]. Представляется целесообразным исследование этого нового подхода с точки зрения применимости к тем или иным проблемам и явлениям. Также необходима разработка методики оценки интересующих величин с необходимой точностью. В одной из первых работ по применению машинного обучения к исследованию фазовых переходов второго рода была предложена методика обучения с учителем [2] и получены удовлетворительные оценки критической температуры модельных систем. Этот подход был дополнен систематическим методом оценки критического показателя корреляционной длины [3]. Подход основан на бинарной классификации с применением глубокой нейронной сети.

Мы поднимаем вопрос об универсальности применения обученной сети. Обучение нейронной сети достаточно затратный процесс как по формированию набора данных для обучения, так и по затра-

там компьютерного времени и компьютерных мощностей. Преимущество нейронные сети имеют при использовании уже обученной сети, которая применяется, например, для высокоскоростной фильтрации данных с Большого адронного коллайдера [4]. Целесообразно изучить универсальность применения нейронной сети для извлечения информации о фазовом переходе второго рода. Один из первых шагов в этом направлении – это исследование применимости нейронной сети, обученной на стандартной модели статистической механики, двумерной модели Изинга, для исследований других моделей. Для простоты постановки тестирование проводилось на модели из того же класса универсальности, но с диагональной анизотропией, которая, как известно, нетривиально влияет на исследуемые свойства [5, 6].

Глубокая нейронная сеть (deep neural network, DNN) [7] – это многослойная сеть, первый слой которой представляет собой матрицу $L \times L$. Слои могут быть полносвязными или сверточными, каждая связь между слоями имеет вес [1], который в процессе обучения изменяется. Особенностью DNN является механизм обратного распространения ошибки, который позволяет эффективно обучать сеть [8] за счет изменения весов. В настоящей работе мы использу-

¹⁾e-mail: lev@landau.ac.ru

ем сверточную нейронную сеть (convolutional neural network, CNN). Читатель может найти детали в [3].

При *обучении* на вход нейронной сети подается распределение спинов на решетке $L \times L$ и указывается принадлежность каждого моментального снимка к ферромагнитной или парамагнитной фазе относительно известной температуры фазового перехода. При *тестировании* обученная сеть получает на вход моментальный снимок спинов решетки $L \times L$ при известной температуре T . На выходе одного нейрона сеть выдает $p_i \in [0, 1]$ как оценку принадлежности тестируемого образца парамагнитной фазе. Другой нейрон выдает принадлежность ферромагнитной фазе $f_i \equiv 1 - p_i$. Поэтому в дальнейшем мы будем анализировать только принадлежность парамагнитной фазе. Повторив тестирование N раз при заданных температуре образцов T и размере решетки L , мы получаем оценку $P(T; L)$ вероятности того, что N образцов, полученных при моделировании при температуре T , принадлежат парамагнитной фазе. Расчет этой функции производится по формуле (4). Из набора функций $P(T; L)$ будет определяться температура фазового перехода с помощью конечномерного анализа, как было предложено в работе [2]. Кроме того, мы исследуем вторые моменты распределения вероятностей принадлежности парамагнитной фазе $D(T; L)$, формула (5). Исследование ширины и полуширины гауссовой функции, приближающей распределение $D(T; L)$, позволяет оценить критический показатель корреляционной длины [3].

2. Исследуемые модели. Гамильтониан обобщенной модели Изинга имеет вид (рис. 1):

$$\mathcal{H} = - \sum_{x,y=1}^L \sigma_{x,y} [J_h \sigma_{x+1,y} + J_v \sigma_{x,y+1} + J_d \sigma_{x+1,y+1}]. \quad (1)$$

Температура фазового перехода этой модели известна из аналитического решения [9]:

$$\begin{aligned} & \sinh \frac{2J_v}{k_B T_c} \sinh \frac{2J_h}{k_B T_c} + \sinh \frac{2J_h}{k_B T_c} \sinh \frac{2J_d}{k_B T_c} + \\ & + \sinh \frac{2J_d}{k_B T_c} \sinh \frac{2J_v}{k_B T_c} = 1, \\ & J_v + J_h > 0, \quad J_h + J_d > 0, \quad J_d + J_v > 0. \end{aligned} \quad (2)$$

Обучение будем проводить на изотропной модели при значении параметров $J_d = 0$, $J_h = J_v = 1$. Тестирование будем проводить для случая диагональной анизотропии с набором параметров $J_h = J_v = J$, $J_d \neq 0$. При этом константа J_d может быть как положительной, так и отрицательной, ограниченной неравенством $J_d + J > 0$.

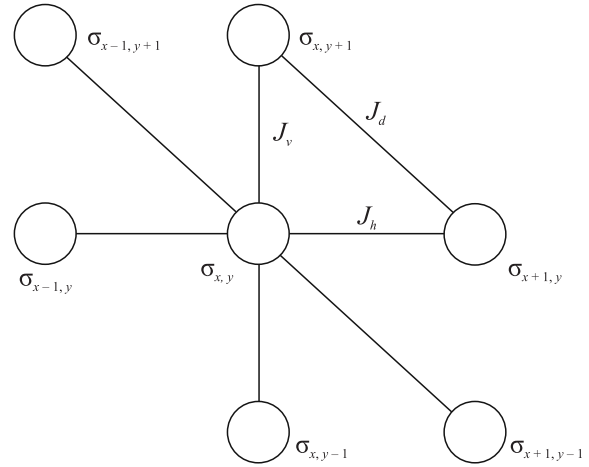


Рис. 1. Часть решетки с узлом $\sigma_{x,y}$ и его окружением в обобщенной модели Изинга

При заданном наборе значений температуры алгоритмом Метрополиса генерируется выборка из N некоррелированных конфигураций спинов. Время термализации составляет $20 \times L^{2.15}$ по оценке [10], и конфигурации сохраняются (*моментальный снимок!*) раз в $2 \times L^{2.15}$ шагов Монте-Карло, измеряемых в числе L^2 локальных переворотов спинов решетки $L \times L$.

3. Обучение нейронной сети. Мы использовали сверточную архитектуру нейронной сети (CNN) с одним сверточным, двумя полносвязными слоями и ReLU-активацией между ними [3]. Сеть обучалась на некоррелированных снимках, полученных при моделировании изотропной модели Изинга, $J_d = 0$, $J_h = J_v = 1$, с линейными размерами решеток $L = 20, 30, 60, 80, 100, 120$. При каждом размере L мы выбирали 100 значений температуры с шагом 0.006 из диапазона температур:

$$T \in [T_c - 0.3, T_c + 0.3], \quad (3)$$

где критическая температура определялась по соотношению (2). Число моментальных снимков при обучении при каждой температуре T и каждом размере L было выбрано $M = 2048$.

Сеть обучалась для каждого линейного размера решетки L одну эпоху во избежание переобучения сети. При количестве эпох больше 20 переобучение ведет к практически однозначной классификации на ферромагнитную и парамагнитную фазы с достаточно точной оценкой температуры перехода, но при этом полностью теряется информация о свойствах корреляционной длины [11].

4. Тестирование. Тестирование проводилось по $N = 512$ моментальным снимкам, полученным при

моделировании для заданных констант связи, при каждом значении T и L , при этом значения L те же, что и при обучении. Значения температуры при тестировании выбирались также из диапазона (3) относительно критической температуры, рассчитанной по формуле (2). Такой выбор естественным образом включает интересующий интервал температур в окрестности фазового перехода тестируемой модели.

Как отмечалось во введении, каждый из N тестируемых образцов получает оценку p_i принадлежности парамагнитной фазе. Для каждого значения температуры и для каждого линейного размера решетки мы рассчитываем оценку вероятности принадлежности образца парамагнитной фазе

$$P(T; L) = \frac{1}{N} \sum_{i=1}^N p_i(T; L) \quad (4)$$

и соответствующие стандартные отклонения

$$D(T; L) = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i(T; L))^2 - \left(\frac{1}{N} \sum_{i=1}^N p_i(T; L) \right)^2}. \quad (5)$$

5. Оценка критической температуры. В работе [2] был предложен метод оценки критической температуры на основе анализа $P(T; L)$ на основании гипотезы о соответствии псевдокритической температуры T_c^* значению оценки вероятности $P(T_c^*; L) = 0.5$. На рисунке 2 приведен результат вычисления набора функций $P(T; L)$ с антиферромагнитным взаимодействием по диагонали $J_d < 0$ при значении параметра анизотропии $J_d/J = -0.5$. Соответствующие функции $D(T; L)$ представлены на рис. 3.

Мы использовали следующие усредненные оценки вероятностей $P(T; L)$ для получения оценки критической температуры:

1. Температура, при которой прогноз сети наиболее близок к 0.5:

$$T_c^*(L) = \min_T |P(T; L) - 0.5|. \quad (6)$$

2. Температура, соответствующая точке пересечения прямой, соединяющей два ближайших к 0.5 прогноза сети, с $y=0.5$:

$$\begin{aligned} T_{\min} &= \min_T |P(T; L) - 0.5|, \quad P(T; L) < 0.5, \\ T_{\max} &= \min_T |P(T; L) - 0.5|, \quad P(T; L) > 0.5, \\ P_{\min} &= P(T_{\min}; L); \quad P_{\max} = P(T_{\max}; L), \\ T_c^{\circ}(L) &= \frac{T_{\min}(P_{\max} - 0.5) + T_{\max}(0.5 - P_{\min})}{P_{\max} - P_{\min}} \end{aligned} \quad (7)$$

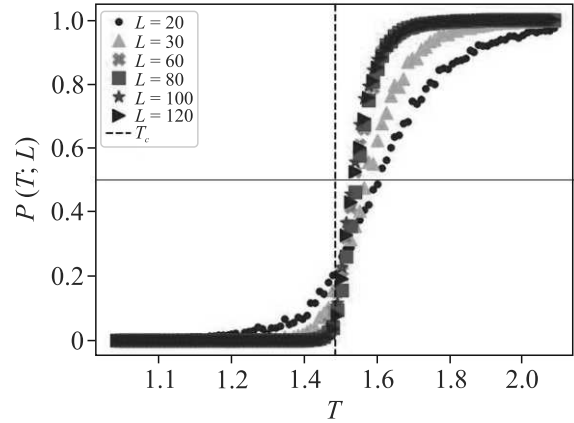


Рис. 2. (Цветной онлайн) Оценка принадлежности парамагнитной фазе $P(T; L)$ (4) модели Изинга с диагональной анизотропией при значении параметра анизотропии $J_d/J = -0.5$

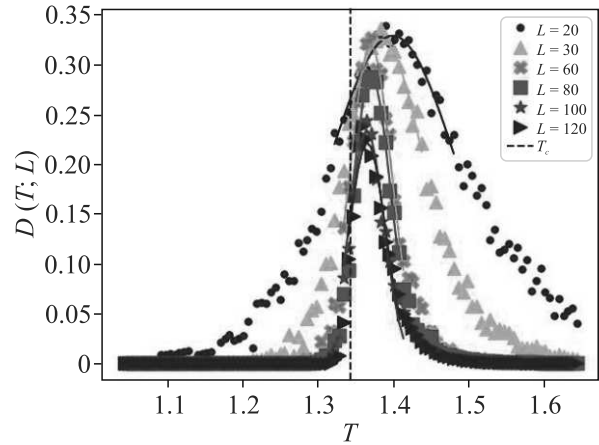


Рис. 3. (Цветной онлайн) Функция стандартных отклонений принадлежности парамагнитной фазе $D(T; L)$ модели Изинга с диагональной анизотропией, $J_d/J = -0.5$

3. Температура на основе функции $D(T; L)$: распределение $D(T; L)$ при каждом L приближается гауссовой функцией со средним μ и стандартным отклонением σ . Значение $\mu(L)$ рассматриваем как оценку критической температуры.

Для каждого набора мы рассчитали асимптотическое значение оценки критической температуры, используя известное выражение для конечномерного сдвига псевдокритической температуры [12, 13]. Например, для оценки по варианту 1, это можно записать как

$$T_c^* = T_c^*(L) + a/L. \quad (8)$$

Отклонение полученной линейной аппроксимации точек $T_c^*(L)$ относительно точной критической температуры представлено на рис. 4 черными квадратами вместе со статистической ошибкой. Аналогично рассчитываются оценки критической температуры для остальных двух случаев. Заметим, что все три способа оценки критической температуры дают статистически эквивалентные результаты. Отклонение оценки критической температуры от точно известного значения обсуждается в разделе 7 статьи.

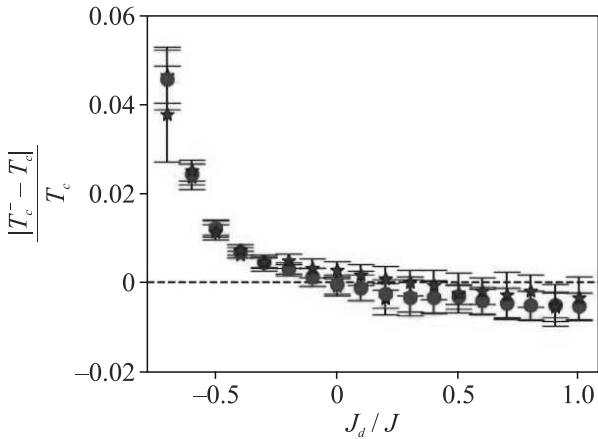


Рис. 4. (Цветной онлайн) Относительная ошибка оценки критической температуры T_c^- в пределе бесконечного размера решетки как функция параметра анизотропии J_d/J . Символы соответствуют трем способам оценки псевдокритической температуры, указанным в тексте: черные квадраты — T_c^- ; зеленые круги — T_c^* ; синие звезды — μ . J_d/J

6. Оценка критического показателя корреляционной длины. Корреляционная длина ξ в модели Изинга (1) в области фазового перехода растет по степенному закону $\xi \propto \tau^{-1/\nu}$ при уменьшении приведенной температуры $\tau = (T - T_c)/T_c$. Критический показатель ν равен единице [14].

При фиксированном L каждую функцию $D(T; L)$ мы приближаем гауссианом, отбрасывая статистически незначимые значения на краях температурного диапазона, и анализируем полученные значения $\mu(L)$ так, как было указано в пункте 3 предыдущего раздела для оценки критической температуры, а также ширину $\sigma(L)$. Здесь мы используем гипотезу [3], о том, что ширина $\sigma(L)$ масштабируется так же, как и некоторые термодинамические функции [12, 13], с показателем корреляционной длины ν :

$$\sigma(L) \propto \frac{b}{L^{1/\nu}}. \quad (9)$$

Эта гипотеза была подтверждена в статье [3] для моделей в двух классах универсальности, в которой

были получены достоверные оценки показателя корреляционной длины для двумерной модели Изинга на квадратной решетке и модели Бакстера–Ву, формулируемой на треугольной решетке.

Результаты оценки критического показателя корреляционной длины изображены на рис. 5, построенном по данным в табл. 1.

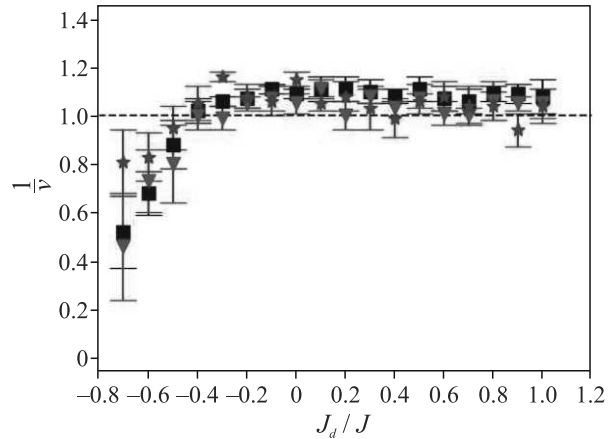


Рис. 5. (Цветной онлайн) Оценка обратного значения критического показателя корреляционной длины: черные квадраты — оценка по полной ширине, красные треугольники по правой части, зеленые звезды — по левой части

Таблица 1. Оценки обратного значения критического показателя корреляционной длины ν , полученные из анализа ширины $\sigma(L)$ функции $D(T; L)$ — второй столбец, правой полуширины $\sigma_r(L)$ этой функции — третий столбец и левой полуширины $\sigma_l(L)$ этой функции — четвертый столбец

J_d/J	$\frac{1}{\nu}$	$(\frac{1}{\nu})_r$	$(\frac{1}{\nu})_l$
-0.7	0.52(15)	0.46(22)	0.81(13)
-0.6	0.68(9)	0.73(13)	0.83(10)
-0.5	0.88(10)	0.80(16)	0.95(9)
-0.4	1.02(5)	1.00(6)	1.05(7)
-0.3	1.06(2)	0.99(5)	1.16(2)
-0.2	1.07(4)	1.05(3)	1.08(5)
-0.1	1.11(2)	1.07(5)	1.06(6)
0.0	1.09(2)	1.05(4)	1.15(3)
0.1	1.11(5)	1.11(4)	1.05(3)
0.2	1.11(5)	1.00(6)	1.08(6)
0.3	1.10(5)	1.08(3)	1.03(9)
0.4	1.08(3)	1.03(3)	0.99(8)
0.5	1.11(5)	1.07(6)	1.06(3)
0.6	1.07(5)	1.01(5)	1.08(6)
0.7	1.06(5)	1.00(3)	1.04(8)
0.8	1.09(3)	1.06(8)	1.04(6)
0.9	1.09(4)	1.05(3)	0.94(7)
1.0	1.08(7)	1.05(6)	1.04(7)

7. Обсуждение результатов. Мы использовали сверточную нейронную сеть для решения задачи бинарной классификации *изотропной* модели Изинга. По анализу предсказаний нейронной сети мы определили температуру фазового перехода T_c и показателя корреляционной длины ν с хорошим совпадением [2, 3] с точным аналитическим результатом [14].

Обученную сеть мы использовали для тестирования модели Изинга с *диагональной анизотропией* с ферромагнитными и антиферромагнитными связями. В области умеренных значений показателя анизотропии обученная на изотропной модели сеть предсказывает температуру фазового перехода и значение показателя корреляционной длины в пределах статистической погрешности, что можно видеть из табл. 1 и рис. 4 и 5 при значениях параметра анизотропии J_d/J от ферромагнитной анизотропии при $J_d/J = 1$ до антиферромагнитной анизотропии при $J_d/J = -0.4$.

При значениях параметра анизотропии $J_d/J < -0.4$ наблюдается систематическое отклонение как температуры перехода, так и показателя корреляционной длины от точных значений. Это отклонение можно объяснить особенностью фазовой диаграммы модели, исследованной Стефенсоном [15] и наличием в тестируемом температурном интервале моментальных снимков распределения спинов, попадающих выше температуры, которая в статье [15] была названа *disorder temperature* T_D , выше которой имеется область с осциллирующими парными корреляциями спинов. Осцилляции обусловлены конкурирующим диагональным антиферромагнитным взаимодействием. Поскольку такие осцилляции полностью отсутствовали при обучении нейронной сети, то результаты тестирования приводят к отклонению оценок критической температуры и показателя корреляционной длины.

Заметим, что диагональная анизотропия нетривиальным образом влияет также на наблюдаемые термодинамические величины. Например, было предсказано аналитически [16] и обнаружено численно [17], что величина кумулянта Биндера, построенная из отношения квадрата второго момента намагниченности к четвертому моменту намагниченности, изменяется в зависимости от параметра диагональной анизотропии. Ранее считалось, что его значение в точке фазового перехода определяется исключительно классом универсальности модели. Вариация кумулянта Биндера при положительных значениях параметра анизотропии невелика и составляет менее процента, как показано на рис. 4

из статьи [18]. При этом можно увидеть резкое падение при значении параметра анизотропии -0.4 , аналогично нашим данным по перекрестному обучению/тестированию.

Таким образом, мы установили качественно границы применимости перекрестного обучения при тестировании нейронной сетью модели с диагональной анизотропией и антиферромагнитным взаимодействием. Эти границы связаны с особенностями фазовой диаграммы анизотропной модели [15] и близостью к точке перехода системы в полностью фрустрированное состояние при $J_d/J = -1$, вблизи которого существуют долгоживущие топологические возбуждения, изученные в работах [19, 20].

Финансирование работы. Работа поддержана грантом 22-11-00259 Российского научного фонда.

Для проведения компьютерного моделирования и машинного обучения использован суперкомпьютерный комплекс Национального исследовательского университета “Высшая школа экономики”.

Конфликт интересов. Авторы данной работы заявляют, что у них нет конфликта интересов.

1. G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, *Rev. Mod. Phys.* **91**, 045002 (2019).
2. J. Carrasquilla and R.G. Melko, *Nat. Phys.* **13**, 431 (2017).
3. V. Chertentkov, E. Burovski, and L. Shchur, *Phys. Rev. E* **108**, L031102 (2023).
4. D. Derkach, M. Hushchyn, T. Likhomanenko, A. Rogozhnikov, N. Kazeev, V. Chekalina, R. Neychev, S. Kirillov, F. Ratnikov, and on behalf of the LHCb collaboration, *J. Phys.: Conf. Ser.* **1085**, 042038 (2018).
5. V. Dohm and S. Wessel, *Phys. Rev. Lett.* **126**, 060601 (2021).
6. V. Dohm, *Phys. Rev. E* **108**, 044149 (2023).
7. R. Kumari and S.K. Srivastava, *International Journal of Computer Applications* **160**, 11 (2017).
8. D.E. Rumelhart, G.E. Hinton, and R.J. Williams, *Nature* **323**(6088), 533 (1986).
9. R. M. F. Houtappel, *Physica* **16**, 425 (1950).
10. V. Chertentkov and L. Shchur, *J. Phys. Conf. Ser.* **740**, 012003 (2021).
11. В. И. Чертенков, кандидатская диссертация, НИУ ВШЭ (2024).
12. M. E. Fisher and A. E. Ferdinand, *Phys. Rev. Lett.* **19**, 169 (1967).
13. A. E. Ferdinand and M. E. Fisher, *Phys. Rev. B* **185**, 832 (1969).
14. L. Onsager, *Phys. Rev.* **65**, 117 (1941).

15. J. Stephenson, Phys. Rev. B **1**, 4405 (1970).
16. X.S. Chen and V. Dohm, Phys. Rev. E **70**, 056136 (2005).
17. W. Selke and L.N. Shchur, J. Phys. A: Math. Gen. **38**, L739 (2005).
18. W. Selke and L.N. Shchur, Phys. Rev. E **80**, 042104 (2009).
19. S.E. Korshunov, Phys. Rev. B **72**, 144417 (2005).
20. A. Smerald, S. Korshunov, and F. Mila, Phys. Rev. Lett. **116**, 197201 (2016).