УДК 519.6

**Байесовские нейронные сети для регрессии и экстраполяции результатов no-core shell model**

**Р.Э. Шарыпов**

*Тихоокеанский государственный университет (г. Хабаровск)*

*2017104939@togudv.ru*

*В работе рассматривается применение байесовских нейронных сетей с обучаемыми априорными распределениями для задачи регрессии и экстраполяции результатов расчётов в рамках no-core shell model (NCSM). NCSM обеспечивает микроскопически точное описание структуры ядер, однако его вычислительная сложность резко возрастает с увеличением числа нуклонов, что делает прямые расчёты невозможными в ряде областей. Разработанная модель использует трёхслойную архитектуру с вариационными стохастическими слоями, обучаемыми с помощью вариационного вывода в библиотеке TensorFlow Probability. В процессе обучения применяются методы повышения стабильности и точности: механизм ранней остановки, косинусный режим изменения скорости обучения с фазой разогрева (warmup) и ограничение нормы градиентов. Байесовский подход обеспечивает не только высокую точность аппроксимации результатов NCSM, но и количественную оценку неопределённости предсказаний, что критически важно при экстраполяции в область, недоступную для прямых вычислений.*

No-core shell model (NCSM) даёт микроскопически обоснованные значения энергий и других характеристик лёгких ядер, но его применение ограничено из-за экспоненциального роста вычислительных затрат при увеличении размерности модельного пространства [1, 2]. В данной работе рассматривается построение вероятностной аппроксимирующей модели, основанной на байесовских нейронных сетях (Bayesian Neural Networks, BNN), для приближения и частичной экстраполяции результатов NCSM с оценкой апостериорной неопределённости предсказаний [3, 4].

Модель представляет собой компактную многослойную нейронную сеть с вариационными (стохастическими) слоями, реализованную с помощью TensorFlow Probability [5, 6]. Для параметров сети используются нормальные априорные распределения; постериорные приближения для весов задаются в вариационной форме (также нормальными). Целевая функция — отрицательное лог-правдоподобие наблюдаемых значений на фоне регуляризации в виде дивергенции Кульбака–Лейблера (ELBO):

 (1)

где первый член – среднее (по вариационному распределению параметров) отрицательное лог-правдоподобие наблюдений (negative log-likelihood); второй – регуляризационный член: взвешенная с помощью параметра дивергенция Кульбака–Лейблера между вариационным приближением и априорным распределением параметров. Данные: – входные признаки – соответствующие целевые значения (вектор наблюдений). Реализация допускает как фиксированные, так и обучаемые априорные расределения; последний вариант требует дополнительных мер стабилизации из-за роста числа параметров.

В докладе сравниваются нейронные сети с тремя различными нормальными априорными распределениями: фиксированным, с обучаемыми и с параметрами , унаследованными от подхода с использованием ансамбля детерминированных нейронных сетей (как в работе [7]).

Исходные данные (рис. 1) представляют собой результаты расчетов в NCSM и зависят от двух параметров – размера базисного пространства модели, определяемого максимальным числом осцилляторных квантов возбуждения и величины осцилляторной энергии . Они предварительно масштабируются методом MinMax. Масштабирующие функции сохраняются и используются при выводе для обратного приведения предсказаний к физическим единицам. Данные разбиваются на обучающую, валидационную и тестовую выборки. Также проводится их интерполяция с помощью кубических сплайнов.



*Рис. 1.* Исходные данные. Зависимость от ℏΩ энергии основного состояния ядра

6He в расчетах с NN-потенциалом Daejeon16.

Для стабилизации процесса обучения применяется косинусный режим скорости обучения с фазой разогрева (cosine decay with warmup), который представляет собой устойчивую и широко используемую технику оптимизации нейросетевых моделей. Идея состоит в двух последовательных этапах: (i) *warmup* — кратковременное плавное увеличение шага оптимизации от малого начального значения до целевого значения, что стабилизирует начальную фазу адаптивных оптимизаторов (таких как Adam) и уменьшает риск больших начальных градиентов; (ii) *cosine decay* — последующее монотонное снижение шага по косинусному закону, обеспечивающее гладкое уменьшение величины обновлений и способствующее тщательной донастройке в поздней фазе обучения. В практическом исполнении длину warmup выбирают либо в абсолютных шагах, а минимальный уровень learning rate фиксируют строго положительным (или используют «пороговую» величину), чтобы избежать полного «замирания» обучения. Косинусное затухание может применяться в варианте с перезапусками, где после снижения до малого значения скорость периодически восстанавливается, либо в монотонном варианте без рестартов, как в настоящей работе; выбор зависит от задачи и желаемого поведения оптимизации. На практике сочетание warmup + cosine часто сопровождают дополнительными приёмами: линейным или экспоненциальным снижением веса KL (KL-annealing) при вариационном обучении, усечением градиентов, а также увеличением параметра в Adam для повышения численной устойчивости.

Режим скорости обучения показан на рис. 2. Для предотвращения взрывного роста градиентов применяется усечение (gradient clipping): Усечение по норме (clipnorm) действует глобально на все обучаемые параметры, включая параметры обучаемого априорного распределения, поэтому оно эффективно и для тех случаев, когда априорное распределение является обучаемым.

Ключевой механизм контроля длительности обучения – мониторинг целевой функции на валидационной выборке. Этот механизм останавливает обучение при отсутствии улучшения метрики в течение некоторого окна эпох и восстанавливает веса из лучшего найденного чекпоинта.



*Рис. 2.* Режим скорости обучения.

Использование BNN для аппроксимации и последующей экстраполяции расчетов NCSM является перспективным: модель с вероятностной формулировкой даёт интерпретируемую оценку неопределённости и в ряде случаев пригодна для экстраполяции в недоступные зоны NCSM. Вместе с тем trainable prior требует осторожной настройки (инициализация, регуляризация, контроль числа параметров и оптимизация), особенно при ограниченной обучающей выборке; предложенные практики существенно повышают устойчивость и качество предсказаний.

**Л И Т Е Р А Т У Р А**

1. Navrátil P., Vary J.P., Barrett B.R. Large-basis ab initio no-core shell model and its application to ¹²C // Phys. Rev. C. – 2000. – Vol. 62. – Article 054311.
2. Navrátil P., Quaglioni S., Stetcu I., Barrett B.R. Recent developments in the no-core shell model // arXiv preprint arXiv:0904.0463, 2009.
3. Neal R.M. Bayesian Learning for Neural Networks. // New York: Springer-Verlag, 1996. – Lecture Notes in Statistics.
4. Blundell C., Cornebise J., Kavukcuoglu K., Wierstra D. Weight Uncertainty in Neural Networks // Proc. 32nd Int. Conf. on Machine Learning (ICML). – 2015. – arXiv:1505.05424.
5. TensorFlow Probability. URL: https://www.tensorflow.org/probability (дата обращения: 10.08.2025).
6. Dillon J.V., Langmore I., Tran D., Brevdo E., Vasudevan S., Moore D., Patton A., Alemi A., Hoffman M., Saurous R. TensorFlow Distributions // arXiv preprint arXiv:1711.10604, 2017.
7. R.E. Sharypov, A.I. Mazur, A.M. Shirokov, Machine learning for extrapolating no-core shell model results to infinite basis // Journal of Subatomic Particles and Cosmology, – 2025, V. 4, 100083
8. Loshchilov I., Hutter F. SGDR: Stochastic Gradient Descent with Warm Restarts // arXiv preprint arXiv:1608.03983, 2016.
9. Goyal P., Dollár P., Girshick R., Noordhuis P., Wesolowski L., Kyrola A., Tulloch A., Jia Y., He K. Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour // arXiv preprint arXiv:1706.02677, 2017.
10. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. Attention is All You Need // Advances in Neural Information Processing Systems (NIPS). – 2017. – P. 6000–601