Physics Letters B 778 (2018) 48–53

Contents lists available at ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

Nuclear mass predictions based on Bayesian neural network approach with pairing and shell effects

Z.M. Niu (牛中明)[a,b], H.Z. Liang (梁豪兆)[b,c,d,*]

[a] School of Physics and Materials Science, Anhui University, Hefei 230601, China
[b] Interdisciplinary Theoretical Science Research Group, RIKEN, Wako 351-0198, Japan
[c] RIKEN Nishina Center, Wako 351-0198, Japan
[d] Department of Physics, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan

ARTICLE INFO

Article history:
Received 14 August 2017
Received in revised form 30 December 2017
Accepted 2 January 2018
Available online 9 January 2018
Editor: W. Haxton

ABSTRACT

Bayesian neural network (BNN) approach is employed to improve the nuclear mass predictions of various models. It is found that the noise error in the likelihood function plays an important role in the predictive performance of the BNN approach. By including a distribution for the noise error, an appropriate value can be found automatically in the sampling process, which optimizes the nuclear mass predictions. Furthermore, two quantities related to nuclear pairing and shell effects are added to the input layer in addition to the proton and mass numbers. As a result, the theoretical accuracies are significantly improved not only for nuclear masses but also for single-nucleon separation energies. Due to the inclusion of the shell effect, in the unknown region, the BNN approach predicts a similar shell-correction structure to that in the known region, e.g., the predictions of underestimation of nuclear mass around the magic numbers in the relativistic mean-field model. This manifests that better predictive performance can be achieved if more physical features are included in the BNN approach.

Mass is a fundamental property of atomic nuclei. It can be employed to extract various nuclear structure information, such as nuclear pairing correlation, shell effect, deformation transition, and so on [1]. Nowadays it has been also widely used to determine nuclear effective interactions [2]. Moreover, nuclear mass is essential to determine the nuclear reaction energy in astrophysics and hence plays a crucial role in understanding the origin of elements in Universe [3]. In addition, the accurate mass determination is very important to test the unitarity of Cabibbo–Kobayashi–Maskawa matrix [4,5].

Measurements of nuclear mass have achieved great progress in recent years [6,7] and about 3000 nuclear masses have been measured up to now [8]. However, the accurate predictions of nuclear mass are still a great challenge for theoretical models, due to the difficulties in the exact theory of nuclear interaction and in the quantum many-body calculations. Nowadays three types of nuclear models are mainly used in global mass predictions: macroscopic, macroscopic–microscopic, and microscopic mass models. The Bethe–Weizsäcker (BW) mass formula is the first model used to estimate nuclear masses [9,10], which belongs to the macroscopic type. It assumes the nucleus is similar to a charged liquid drop, so the microscopic effects, such as shell effect, cannot be well described. By taking into account the important corrections related to the microscopic effects, the macroscopic–microscopic models are developed, such as the finite-range droplet model (FRDM) [11] and the Weizsäcker–Skyrme (WS) model [12]. The microscopic mass models are mainly rooted in the density functional theory, which are more complicated but potentially have a better ability of extrapolation. In the non-relativistic framework, a series of Hartree–Fock–Bogoliubov (HFB) mass models have been constructed with the Skyrme [13,14] or Gogny [15] effective interactions. In recent years, the relativistic mean-field (RMF) model also receives wide attention due to its success in describing various nuclear phenomena [16–22] and its successful applications in astrophysics [23–26]. Based on the RMF model, global calculations of nuclear mass have been made and the accuracies were gradually improved [27–29].

The accuracies of these mass models range from about 3 MeV for the BW model [30] to about 0.3 MeV for the WS model [12]. However, these accuracies are still insufficient to the studies of exotic nuclear structures and astrophysics nucleosynthesis. Especially, these models predict very different nuclear masses with the dif-

* Corresponding author.
E-mail addresses: zmniu@ahu.edu.cn (Z.M. Niu), haozhao.liang@riken.jp (H.Z. Liang).

https://doi.org/10.1016/j.physletb.2018.01.002

ferences even up to tens of MeV when they are extrapolated to the neutron drip line. Therefore, it is still a high demand to further improve the existing nuclear mass models. Some techniques have been developed along this direction, such as the radial basis function (RBF) approach [31–34] and the image reconstruction technique with the CLEAN algorithm [35]. Moreover, the neural network has been proved to be a very powerful tool and it has been widely used in an impressive range of problem domains, such as pattern recognition and machine learning, see, e.g., books [36,37] and the references therein. The application of neural network to predict nuclear masses can be traced back to the 1990s [38]. A series of works after that were developed to further improve its predictive performance [39–41]. It was also extended to study other nuclear properties, such as nuclear β-decay half-lives [42]. These approaches usually need many parameters, in general hundreds or even thousands of parameters, for achieving better predictions, so the over-fitting problem and the quantification of uncertainties in the predictions should be treated in a reliable way.

The Bayesian approach can avoid the over-fitting problem by introducing the prior distribution of parameters, and it can quantify the uncertainties in the predictions since all parameters have probability distributions [43]. Thus, it would be a valuable approach for improving the mass predictions of nuclear models. However, the Bayesian approach involves high-dimensional integrals over the whole parameter space, so its calculations are very time-consuming and great progress was achieved only in the last decades along with the developments in sampling methods and dramatic improvements in the speed and memory of computers [37]. Recently, the Bayesian neural network (BNN) approach was applied to improve the theoretical predictions of nuclear masses [44] and nuclear charge radii [45]. The noise error in the likelihood function is a key quantity in the BNN approach, however, it was usually much simplified by taking a fixed value in the previous studies [44,45]. In this work, we will introduce a prior distribution for the noise error. Furthermore, only the proton and mass numbers were considered in the input layer of the neural network in the previous studies [44,45]. Here we will consider more physical features into the input layer, i.e., we will include two quantities related to the well known nuclear pairing and shell effects, and investigate their influences on the predictive performance of the BNN approach.

In the Bayesian approach, the model parameters $\boldsymbol{\omega}$ are described probabilistically. A probability distribution $p(\boldsymbol{\omega})$ is introduced over all possible values of $\boldsymbol{\omega}$ based on our background knowledge, which is called the *prior* distribution. When we observe a set of data $D = \{(\boldsymbol{x_1}, t_1), (\boldsymbol{x_2}, t_2), ..., (\boldsymbol{x_N}, t_N)\}$, this distribution will be updated by using the Bayes' theorem

$$p(\boldsymbol{\omega}|D) = \frac{p(D|\boldsymbol{\omega})p(\boldsymbol{\omega})}{p(D)} \propto p(D|\boldsymbol{\omega})p(\boldsymbol{\omega}), \tag{1}$$

where $\boldsymbol{x_n}$ and t_n $(n = 1, 2, ..., N)$ are input and output data, N is the number of data; $p(D|\boldsymbol{\omega})$ is the likelihood function, which contains the information about parameters $\boldsymbol{\omega}$ derived from the observations; $p(\boldsymbol{\omega}|D)$ is the probability distribution of parameters $\boldsymbol{\omega}$ after the data D are considered, which is called the *posterior* distribution; $p(D)$ is a normalization constant, which ensures the posterior distribution is a valid probability density and integrates to one.

For the likelihood function $p(D|\boldsymbol{\omega})$, a Gaussian distribution, $p(D|\boldsymbol{\omega}) = \exp(-\chi^2/2)$, is usually employed, where the objective function χ^2 reads

$$\chi^2 = \sum_{n=1}^{N} \left(\frac{t_n - S(\boldsymbol{x}; \boldsymbol{\omega})}{\Delta t_n} \right)^2. \tag{2}$$

Here, the standard deviation parameter Δt_n is the associated noise error related to the nth observable. For the BNN approach, the function $S(\boldsymbol{x}; \boldsymbol{\omega})$ is described with a neural network, which is

$$S(\boldsymbol{x}; \boldsymbol{\omega}) = a + \sum_{j=1}^{H} b_j \tanh\left(c_j + \sum_{i=1}^{I} d_{ji} x_i \right), \tag{3}$$

where $\boldsymbol{x} = \{x_i\}$ and $\boldsymbol{\omega} = \{a, b_j, c_j, d_{ji}\}$, and H and I are the numbers of neurons in the hidden layer and the number of input variables, respectively. In total, the number of parameters in this neural network is $1 + (2 + I) * H$.

For the prior distributions $p(\boldsymbol{\omega})$ of model parameters, they are usually set as Gaussian distributions with zero means. However, the precisions (inverse of variances) of these Gaussian distributions are not set as fixed values by hand. We set them as gamma distributions so that the precisions can vary over a large range and hence the BNN approach can search the optimal values of precisions in the sampling process automatically.

After specifying the likelihood function $p(D|\boldsymbol{\omega})$ and the prior distribution $p(\boldsymbol{\omega})$, the posterior distribution $p(\boldsymbol{\omega}|D)$ of model parameters is known in principle. One can then make predictions based on this posterior distribution,

$$\langle S \rangle = \int S(\boldsymbol{x}; \boldsymbol{\omega}) p(\boldsymbol{\omega}|D) d\boldsymbol{\omega}. \tag{4}$$

Since the model parameters are described with a probability distribution, an estimate of uncertainty in theoretical predictions is obtained naturally as

$$\Delta S = \sqrt{\langle S^2 \rangle - \langle S \rangle^2}. \tag{5}$$

Note that Eq. (4) involves a high-dimensional integral in the whole parameter space. For that, we will employ the Monte Carlo integral algorithm, where the posterior distribution $p(\boldsymbol{\omega}|D)$ is sampled using the flexible Bayesian model developed by Neal [43], in which the Markov chain Monte Carlo algorithm is employed.

In this work, we will employ the BNN approach to reconstruct mass residuals between experimental data $M^{\rm exp}$ and mass predictions $M^{\rm th}$ of various models, i.e.,

$$t_n = M^{\rm exp}(\boldsymbol{x}) - M^{\rm th}(\boldsymbol{x}). \tag{6}$$

As in Refs. [44,45], the inputs are usually taken as $\boldsymbol{x} = (Z, A)$. However, we will consider more physical information into the BNN approach, so two extra inputs δ and P related to nuclear pairing and shell effects are also included, which are

$$\delta = [(-1)^Z + (-1)^N]/2, \quad P = \nu_p \nu_n / (\nu_p + \nu_n). \tag{7}$$

Here, ν_p and ν_n are the differences between the actual nucleon numbers Z and N and the nearest magic numbers (8, 20, 28, 50, 82, 126 for protons and 8, 20, 28, 50, 82, 126, 184 for neutrons) [30]. For simplicity, we will use BNN-I2 and BNN-I4 to denote the BNN approaches with $\boldsymbol{x} = (Z, A)$ and $\boldsymbol{x} = (Z, A, \delta, P)$, respectively. Their numbers of neurons are taken as $H = 42$ and $H = 28$, respectively, so the model parameters in both neural networks are the same as 169.

The experimental masses are taken from the atomic mass evaluation of 2016 (AME2016) [8], while only those nuclei with $Z, N \geqslant 8$ and experimental errors $\sigma^{\rm exp} \leqslant 100$ keV are considered. There are 2272 data left that compose the entire data set. In order to examine the validity of the BNN approach, we separate the entire set into two different sets: the learning set and the validation set. The learning set is built by randomly selecting 1800 nuclei from the entire set and the remaining 472 nuclei compose the