Nuclear masses in extended kernel ridge regression with odd-even effects

X. H. Wu\textsuperscript{a}, L. H. Guo\textsuperscript{a}, P. W. Zhao\textsuperscript{a,}\textsuperscript{*}

\textsuperscript{a}State Key Laboratory of Nuclear Physics and Technology, School of Physics, Peking University, Beijing 100871, China

Abstract

The kernel ridge regression (KRR) approach is extended to include the odd-even effects in nuclear mass predictions by remodulating the kernel function without introducing new weight parameters and inputs in the training network. By taking the WS4 mass model as an example, the mass for each nucleus in the nuclear chart is predicted with the extended KRR network, which is trained with the mass model residuals, i.e., deviations between experimental and calculated masses, of other nuclei with known masses. The resultant root-mean-square mass deviation from the available experimental data for the 2353 nuclei with $Z \geq 8$ and $N \geq 8$ can be reduced to 128 keV, which provides the most precise mass model from machine learning approaches so far. Moreover, the extended KRR approach can avoid the risk of worsening the mass predictions for nuclei at large extrapolation distances, and meanwhile, it provides a smooth extrapolation behavior with respect to the odd and even extrapolation distances.

Keywords: extended kernel ridge regression, nuclear masses, odd-even effects, machine learning

1. Introduction

Nuclear mass is of fundamental importance not only for various aspects of nuclear physics \cite{1, 2}, but also for astrophysics \cite{3, 4}. It can be used to extract a lot of nuclear structure information, e.g., nuclear deformation \cite{5, 6}, shell structure \cite{7, 8}, and effective interactions \cite{9, 10}. It is also a key nuclear physics input in understanding the energy production in stars \cite{11} and the origin of elements in the universe \cite{12} by determining the reaction energies of all the involved nuclear reactions.

Great achievements in nuclear mass measurements have been made during recent decades with the development of radioactive ion beam (RIB) facilities, and about 2500 nuclear masses have been measured to date \cite{13, 14}. Nevertheless, the masses of a large number of neutron-rich nuclei involved in the $r$-process remain unknown from experiments and cannot be measured even with the next-generation RIB facilities. The local mass relations, such as the isobaric multiplet mass equation (IMME) \cite{15}, the Garvey-Kelson (GK) relations \cite{16}, and the residual proton-neutron interactions \cite{17}, can be used to predict the masses of nuclei very close to the experimentally known region, but they are not sufficient for the demands of $r$-process simulations. Therefore, theoretical predictions for nuclear masses are imperative at the present time. It can be traced back to the von Weizsäcker mass formula based on the famous liquid drop model (LDM) \cite{18}. Tremendous efforts have been made in pursuing different possible extensions of the LDM, which are known as the macroscopic-microscopic models, such as the finite-range droplet model (FRDM) \cite{19} and the Weizsäcker-Skyrme (WS) model \cite{20}. The microscopic mass models based on the nonrelativistic and relativistic density functional theories (DFTs) have also been developed (see e.g., Refs. \cite{21, 22, 23, 24, 25, 26} and references therein). They are usually believed to have a better reliability of extrapolation \cite{27}, although their precisions of predicting experimentally known masses are currently poorer than the macroscopic-microscopic models.

The root-mean-square (rms) deviations between theoretical mass models and the available experimental data \cite{12, 13} range from about 3 MeV for the BW model \cite{28} to about 300 keV for the WS ones \cite{19}, which are still not enough for accurate studies of exotic nuclear structure and astrophysical nucleosynthesis. In particular, for neutron-rich nuclei far away from the experimentally known region, the differences among the predictions of different mass models can be as large as several tens MeV. In recent years, machine learning approaches have been employed to further improve the accuracies of nuclear models, such as the radial basis function (RBF) approach \cite{29, 30, 31, 32, 33, 34}, the Bayesian neural network (BNN) approach \cite{35, 36, 37, 38}, and the kernel ridge regression (KRR) approach \cite{39}. By training the machine learning network with the mass model residuals, i.e., deviations between experimental and calculated masses, machine learning approaches can reduce the corresponding rms deviations to about 200 keV. However, the RBF and the BNN approaches predict quite different masses for nuclei far away from the region with known masses \cite{40}. This indicates that the extrapolation abilities of these two machine learning tools have not been properly understood yet. In contrast, the KRR approach with the Gaussian kernel can

\textsuperscript{*}Corresponding Author

Email address: pzzhao@pku.edu.cn (P. W. Zhao.)
automatically identify the limit of the extrapolation distance and avoid the risk of worsening the mass descriptions for nuclei at large extrapolation distances [33].

The rms deviations can be further reduced with more effects being taken into account, e.g., the odd-even effects, pairing effects, and shell effects. The odd-even effects are included in the RBF approach by building and training additional networks [32], while the pairing and shell effects are included in the BNN approach [55] by including additional inputs in the network. The numbers of weight parameters are significantly increased in comparison with the original networks in both approaches. The accuracies of the mass predictions for known nuclei are further improved, but the RBF and the BNN approaches still predict quite different masses for nuclei far away from the known region [10].

In the present work, the KRR approach is extended to include the odd-even effects for nuclear mass predictions. The odd-even effects are included only by remodulating the KRR kernel function. Therefore, the number of the weight parameters is not increased, and the inputs of the network remain in the present extended KRR approach. The hyperparameters in the extended KRR network are optimized by careful cross-validations. The performance and reliability of the extrapolated mass predictions are analyzed in detail.

2. Theoretical framework

The KRR approach is a powerful machine learning approach for nonlinear regression and has been successfully applied in nuclear mass predictions [33]. In order to include the odd-even effects, the KRR function is extended to be

\[ S(x_j) = \sum_{i=1}^{m} K(x_j, x_i)\alpha_i + \sum_{i=1}^{m} K_{oe}(x_j, x_i)\beta_i, \]  

(1)

where \( x_i \equiv (Z_i, N_i) \) are locations of nuclei in the nuclear chart, \( m \) is the number of training nuclei, \( \alpha_i \) and \( \beta_i \) are weights to be determined, \( K(x_j, x_i) \) and \( K_{oe}(x_j, x_i) \) are kernels, which measure the similarity between nuclei. The closer two nuclei on the nuclear chart, the larger similarity they have. This is measured by the Gaussian kernel

\[ K(x_j, x_i) = \exp(-||x_i - x_j||^2/2\sigma^2), \]  

(2)

where \( ||x_i - x_j|| = \sqrt{(Z_i - Z_j)^2 + (N_i - N_j)^2} \) stands for the distance of two nuclei. The kernel \( K_{oe}(x_j, x_i) \) is introduced to enhance the correlations between nuclei that have the same parity of proton and neutron numbers, which reads

\[ K_{oe}(x_j, x_i) = \delta_{oe}(x_j, x_i) \exp(-||x_i - x_j||^2/2\sigma_{oe}^2). \]  

(3)

Here, \( \delta_{oe}(x_j, x_i) = 1 \) for two nuclei have the same parity of proton and neutron numbers, otherwise \( \delta_{oe}(x_j, x_i) = 0 \). The Gaussian widths \( \sigma \) and \( \sigma_{oe} \) measure the length scale on the distance that the kernels affect. The weights \( \alpha_i \) and \( \beta_i \) are determined by minimizing the loss function defined as

\[ L(\alpha, \beta) = \sum_{i=1}^{m} \left[ (S(x_i) - y(x_i))^2 + \alpha_x^2 K\alpha + \lambda_{oe}e^T K_{oe}\beta \right]. \]  

(4)

The first term is the variance between the data \( y(x_i) \) and the KRR prediction \( S(x_i) \), and the second and third terms are regularizers that penalize large weights to reduce the risk of overfitting. The hyperparameters \( \lambda \) and \( \lambda_{oe} \) determine the regularization strength. Minimizing Eq. (4), we can obtain

\[ \beta = \frac{\lambda}{\lambda_{oe}} \alpha, \]  

(5)

\[ \alpha = \left( K + \frac{\lambda}{\lambda_{oe}} K_{oe} + \lambda I \right)^{-1} y. \]  

(6)

With Eq. (5), the extended KRR function (1) can be rewritten as a standard KRR function

\[ S(x_j) = \sum_{i=1}^{m} K'(x_j, x_i)\alpha_i, \]  

(7)

with the remodulating kernel

\[ K'(x_j, x_i) = K(x_j, x_i) + \frac{\lambda}{\lambda_{oe}} K_{oe}(x_j, x_i). \]  

(8)

Here, the weights \( \alpha_i \) are determined by Eq. (5). Note that the number of weight parameters in the present extended KRR approach is the same as the original KRR due to the relation between \( \alpha \) and \( \beta \) in Eq. (5).

3. Numerical details

The extended KRR function (7) is trained to reconstruct the mass residuals, i.e., the deviations \( \Delta M(N, Z) = M_{exp}(N, Z) - M_{th}(N, Z) \) between the experimental data \( M_{exp} \) and predictions \( M_{th} \) for a given mass model. Once the weights \( \alpha_i \) are obtained, the extended KRR function \( S(N, Z) \) can be calculated with Eq. (7) for every nucleus \( (N, Z) \). The predicted mass for a nucleus \( (N, Z) \) is, thus, given by \( M_{EKR}(N, Z) = M_{th}(N, Z) + S(N, Z) \).

In the present work, the 2353 experimental data of masses \( M_{exp} \) for nuclei with \( Z \geq 8 \) and \( N \geq 8 \) are taken from AME2012 [12]. The theoretical masses \( M_{th} \) are taken from the mass table WS4 [13], which is one of the most accurate nuclear mass tables.

The leave-one-out cross-validation is adopted to determine the hyperparameters (\( \sigma, \lambda, \sigma_{oe}, \lambda_{oe} \)). With a given set of hyperparameters (\( \sigma, \lambda, \sigma_{oe}, \lambda_{oe} \)), the mass prediction for each of the 2353 nuclei can be given with the extended KRR network trained on all other 2352 nuclei. The optimized hyperparameters (\( \sigma = 1.25, \lambda = 0.05, \sigma_{oe} = 2.65, \lambda_{oe} = 0.15 \)) are obtained when the rms deviation between experimental and predicted masses of the 2353 nuclei attains its minimum.
4. Results and discussion

Figure 1: (Color online). Mass differences between the experimental data and the predictions of the WS4, the KRR, and the extended KRR models for the (a) even-even, (b) even-odd, (c) odd-even, and (d) odd-odd nuclei.

The mass predictions $M_{\text{EKRR}}$ for every nuclei can be calculated by the extended KRR approach with the determined hyperparameters, in which the network weights $\alpha_i$ are trained on the set consisting of all other nuclei with experimentally known masses.

The mass differences between the extended KRR predictions and the experimental data are shown in Fig. 1 for the even-even (e-e), even-odd (e-o), odd-even (o-e), and odd-odd (o-o) nuclei, in comparison with the ones of the WS4 \cite{19} and the KRR predictions \cite{39}. It can be clearly seen that the predictive power of the WS4 mass model is further improved by the extended KRR approach in comparison with the KRR approach. The significant improvement of the extended KRR approach is mainly due to the consideration of the odd-even effects, which eliminates the staggering behavior of mass deviations with respect to the even and odd numbers of nucleons in the KRR approach \cite{39}. Quantitatively, the rms deviation $\Delta_{\text{rms}} = 298 \text{ keV}$ of 2353 nuclei for the WS4 mass model, is reduced to 199 keV by the KRR approach, and is further reduced to 128 keV by the extended KRR approach with the odd-even effects. This indeed provides, so far, the most precise mass model from machine learning approaches.

The mass deviations between the data and the predictions are relatively larger for light nuclei. This may be due to the fact that there are large individual differences among nuclear masses of light nuclei, which are difficult to be reproduced by global mass models. On the other hand, for the applications of nuclear masses in the $r$-process simulations, the masses of nuclei heavier than iron are important. In Fig. 2(a), the rms deviations $\Delta_{\text{rms}}$ between the experimental data and the predictions of the WS4, the KRR, and the extended KRR models for $Z \leq 26$, $Z > 26$ nuclei, and the whole set of nuclei are shown. The extended KRR approach improves the descriptions of masses

Figure 2: (Color online). (a) The rms deviations $\Delta_{\text{rms}}$ between the experimental data and the predictions of the WS4, the KRR, and the extended KRR models for $Z \leq 26$, $Z > 26$ nuclei, and the whole set of nuclei. The dashed line is drawn for guiding eyes at the rms deviation of 100 keV. (b) The number of nuclei with the corresponding mass deviations from the data $\Delta M$ locating in various slots, such as $0 < \Delta M \leq 100 \text{ keV}$, $100 < \Delta M \leq 200 \text{ keV}$, etc.
for both light and heavy nuclei significantly. Moreover, for the heavy nuclei with $Z > 26$, the rms deviation between the experimental data and the predictions of the extended KRR model is as small as 100 keV, which is reaching the chaos-related unpredictability limit for the calculation of nuclear masses [41, 54].

In Fig. 2(b), it is depicted that the number of nuclei with the corresponding mass deviations from the data $\Delta M$ locating in various slots, such as $0 < \Delta M \leq 100$ keV, $100 < \Delta M \leq 200$ keV, etc. This gives a detailed analyse for the distributions of the mass deviations. For the KRR model, the mass deviations from data are smaller than 200 keV for most nuclei, and the number of these nuclei is 1721, which is about 73% of the whole nuclei set. For the extended KRR model, however, the mass deviations from data are smaller than 100 keV for more than 70% of the nuclei, and they are smaller than 200 keV for more than 90% of the nuclei. This means that there are only 200 nuclei with mass deviations larger than 200 keV, and most of them are light nuclei. To improve the descriptions of masses for these nuclei would be a challenging task for the future.

Figure 3: (Color online). Comparison of the extrapolation power of the extended KRR, the KRR, the RBF and the RBF approach with odd-even effects (RBFoe) [32] for eight test sets with different extrapolation distances (see text for details). (a) The rms deviations $\Delta_{\text{rms}}$ of the calculated masses from the WS4 mass model (filled squares), the extended KRR (filled circles), the KRR (empty circles), the RBF (empty triangles), and the RBFoe extrapolations (filled triangles) with respect to the experimental masses. (b) For a clearer comparison, the rms deviations $\Delta_{\text{rms}}$ are scaled to the values for the WS4 mass model.

To examine the extrapolation power of the extended KRR approach for neutron-rich nuclei, similar to Ref. [39], for each isotopic chain with $Z > 26$, the eight most neutron-rich nuclei are removed out from the training set, and they are classified into eight test sets respectively, corresponding to the different extrapolation distances from the remain training set in the neutron direction.

In Fig. 3(a), the rms deviations $\Delta_{\text{rms}}$ of the calculated masses for the eight test sets from the WS4 mass model, the extended KRR extrapolations, the KRR ones, the RBF ones, and the RBF ones with odd-even effects (RBFoe) [32] with respect to the experimental masses are shown as functions of the extrapolation distance. An even more clear comparison is shown in Fig. 3(b), where the rms deviations $\Delta_{\text{rms}}$ are scaled to the corresponding ones for the WS4 mass model. One can see that all the four approaches improve the mass descriptions of nuclei with the extrapolation distances smaller than 4. The corresponding rms deviations are reduced by up to $\approx 40\%$ in comparison with the WS4 mass model. However, one can see obvious distinct features between the KRR type of approaches and the RBF type of approaches at extrapolation distances larger than 4. The rms deviations of the RBF and the RBFoe extrapolations are larger than the ones of the WS4 mass model, and they increase rapidly with the extrapolation distance. For the KRR and the extended KRR extrapolations, the rms deviations increase slowly with the extrapolation distance and are similar or even smaller than the WS4 ones at large extrapolation distances. The distinct features are due to the different behaviors of the Gaussian kernel used in the KRR type of approaches and the Linear kernel used in the RBF type of approaches, which have been discussed in detail in Ref. [39].

Obvious odd-even staggerings along the extrapolation distance can be seen in the rms deviations for the RBF and KRR predictions, where the odd-even effects have not been taken into account explicitly. By considering the odd-even effects with the RBFoe approach, however, the odd-even staggerings remain, and even with an opposite phase with respect to the RBF approach. This means that the RBFoe approach may overestimate the odd-even effects. On the contrary, for the extended KRR approach, as shown in Fig. 3(b), the rms deviations vary smoothly with the extrapolation distance. This indicates that the odd-even effects have been well handled in the extended KRR approach. Moreover, with the increase of the extrapolation distance, the extended KRR approach converged to the original WS4 mass model. This keeps the key merit of the KRR approach for mass extrapolations, i.e., it avoids worsening the mass descriptions for nuclei at large extrapolation distances.

5. Summary

In summary, the kernel ridge regression approach has been extended to include the odd-even effects in nuclear mass predictions by remodulating the kernel function. The obtained extended KRR approach does not introduce new
weight parameters and inputs in the training network. The resulting rms mass deviation from the experimental data is reduced from 298 keV for the WS4 mass model to 128 keV for the extended KRR approach, which provides the most precise mass model from machine learning approaches so far. For nuclei heavier than iron, in particular, the rms deviation is as small as 100 keV, which is reaching the chaos-related unpredictability limit for the calculation of nuclear masses. Comparative study has been carried out to examine the extrapolation power of the extended KRR approach to the neutron-rich nuclei region. It reveals that the extended KRR approach can avoid the risk of worsening the mass predictions for nuclei at large extrapolation distances, and meanwhile, it provides a smooth extrapolation behavior with respect to the odd and even extrapolation distances.

Acknowledgments

This work was partly supported by the National Key R&D Program of China (Contracts No. 2018YFA0404400 and No. 2017YFE0116700), the National Natural Science Foundation of China (Grants No. 11875075, No. 11935003, No. 12070131001) and No. 2017YFE0116700), the National Natural Science R&D Program of China (Contracts No. 2018YFA0404400 and No. 2017YFE0116700), the National Natural Science Foundation of China (Grants No. 11875075, No. 11935003, No. 12070131001).
