Exotic Hadrons in Machine Learning

Jiahao Liu,1,2 Zhenyu Zhang,1,2 Jifeng Hu,1,2,* and Qian Wang1,2,†

1Guangdong Provincial Key Laboratory of Nuclear Science, Institute of Quantum Matter, South China Normal University, Guangzhou 510006, China
2Guangdong-Hong Kong Joint Laboratory of Quantum Matter, Southern Nuclear Science Computing Center, South China Normal University, Guangzhou 510006, China

(Dated: February 11, 2022)

We analyzed the invariant mass spectrum of near-threshold exotic states for one-channel with deep neural network. It can extract the scattering length and effective range, which would shed light on the nature of given states, from the experimental mass spectrum. As an application, the mass spectrum of the $X(3872)$ and the $T_{cc}^+$ are studied. The obtained scattering lengths, effective ranges and the most relevant thresholds are consistent with those from fitting directly to the experimental data. The network, which provides another way to analyse the experimental data, can also be applied to other one-channel near-threshold exotic candidates.

Introduction

Color confinement property of Quantum chromodynamics (QCD) allows for the existence of any color neutral object. That challenges the conventional quark model, in which hadrons are made of either quark-antiquark (mesons) or three quarks (baryons). Especially, the observed exotic hadrons beyond the conventional configurations provide a way to decode the mystery of hadronization. Up to now, tens of exotic candidates have been reported and studied from various aspects [1–10]. One important feature is that most of them are very close to nearby thresholds, as an indication of a large mixture of continuum [5]. In principal, all the configurations with the same quantum number can mix with each other. However, that which configuration plays an important role, either large size hadronic molecule or compact object, is still a well-established question. The key value is the probability $1 - \lambda^2$ (with $\lambda^2$ the wave function renormalization constant) of finding continuum in a given physical state. A typical example is deuteron for instance deuteron [11, 12] discussed by Weinberg in 1960s. This method has been intensively used for discussing the nature of exotic candidates in both hadronic molecular picture [5] and compact one [13]. The value of $\lambda^2$ is related to scattering length [5]

$$a = -2\frac{1 - \lambda^2}{2 - \lambda^2} \left( \frac{1}{\gamma} \right) + O\left( \frac{1}{\beta} \right)$$ (1)

and effective range

$$r = -\frac{\lambda^2}{1 - \lambda^2} \left( \frac{1}{\gamma} \right) + O\left( \frac{1}{\beta} \right)$$ (2)

of the elastic channel [5] for one-channel case in the low-energy limit, which means that the formulae work in near-threshold energy region. Here $\gamma = \sqrt{2\mu E_B}$ is the binding momentum with reduced mass $\mu$ and binding energy $E_B$. $\frac{1}{\beta}$ is the order of range correction. Here $\lambda^2 = 0$ and $\lambda^2 = 1$ for pure molecule and compact object, respectively. In other words, to the leading order

$$a = -\frac{1}{\gamma}, \quad r = O\left( \frac{1}{\beta} \right)$$ (3)

for pure molecule and

$$a = -O\left( \frac{1}{\beta} \right), \quad r = -\infty$$ (4)

for compact object. As a result, extraction of the scattering length and effective range from experimental data is a direct way to shed light on the nature of interested hadrons. Recent and typical examples are the case of the $X(3872)$ [14–16] and $T_{cc}^+$ [15, 17–19] from both experimental and theoretical sides. This work aims at developing a deep learning network for automatically extracting the scattering length and effective range from experimental data directly. The final goal is to set up a deep learning network implementing multichannel case. As the first step, this work starts from the one-channel case. This method has been successfully applied to the $P_c(3412)$ [20], the $\pi N$ system [21, 22], the nucleon-nucleon system [23, 24], focusing on various facts. For instance, Ref. [20] sets a classifier, instead of extracting scattering length and effective range, of a given state by a bottom-up approach to avoid model dependence.

Physics framework

The expressions of scattering length (Eq. 1) and effective range (Eq. 2) are obtained by matching the effective range expansion (ERE) scattering amplitude

$$T_{NR}(E) = -\frac{2\mu}{1/a + (r/2)k^2 - i k}$$ (5)

to explicit scattering amplitude, where the subscript “NR” indicates the nonrelativistic expression. Here $\mu = \frac{m_1 m_2}{m_1 + m_2}$ and $E$ are the reduced mass and total energy of the two-particle system. Accordingly, $k = \sqrt{2\mu (E - m_1 - m_2)}$ is the three momentum of the scattering particle in the center-of-mass frame. $m_1$ and $m_2$ are the masses of the two particles.

As the line shape of a state is dominant by the elastic $^1$ scattering amplitude [25] once the elastic channel

---

1 Here elastic channel means the channel strongly coupled to the interested state.
FIG. 1. The structure of ResNet-based neural network used in this work.

FIG. 2. Correlations between predicted value and input label value. (a) is for scattering length $a$. (b),(c),(d) are for effective ranges within the regions $[0.49,0.99]$ fm, $[-9.87,-2.47]$ fm, and $[-2.47,-0.49]$ fm, respectively. (e) is for threshold. (f) is for resolution parameter $\sigma$.

is predominant in the production vertex, one can consider that the line shapes is described by $|T_{NR}(E)|^2$ convoluted with a Gaussian function

$$G(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}},$$

where the mean value is set to zero, up to a phase space factor. $\sigma$ is the resolution which depends on the energy resolution of measuring the invariant mass spectrum. Based on this formula, we generate 150000 line shapes for training with the parameters within the re-
Table I. Parameters of the $X(3872)$ from deep learning (the second column) and fit (the third column) to the data directly.

| $X(3872)$ parameters | Deep Learning | Fit          |
|----------------------|--------------|-------------|
| parameter $a$ (fm)   | 8.76 ± 1.75  | 9.95 ± 0.34 |
| parameter $r$ (fm)   | 0.56 ± 0.55  | 0.32 ± 0.08 |
| parameter threshold (MeV) | 3871.30 ± 0.52 | 3871.20 ± 0.01 |
| parameter $\sigma$ (MeV) | 1.20 ± 0.15  | 1.70 ± 0.16 |

The regions of scattering length and effective range allow for both bound and virtual states [26]. The threshold region covers charmonium(-like) energy region and resolution locates usually experimental region.

Training. A multi-layer-perception [27] based ResNet [28] is implemented with PyTorch[31] to regress four parameters $a$, $r$, threshold and $\sigma$ by training the generated dataset. The parameters $a$ and $r$ are simultaneously regressed with a model, and the rest two are individually regressed with another two models as shown in Fig. 1. Three models are built with an identical structure, in which the input layer is set to $200 \times 1$ vector, followed by a dimensionality reduction layer and three ResBlocks, then compressed into one or two outputs, and finally connected to the parameter labels. The ResBlock introduces a shortcut connection between the relu non-linear activation layer and the last layer of the block. In this way, solving the models with the Adam [30] optimizer is of high efficiency if we choose an optimization metric: the mean squared error function. A reasonable solution could be obtained using an initial learning rate value of 0.001, and randomizing the neuron weights with a normal distribution while setting the neuron bias to zero. Note that our labeled values have been applied with normalization and nondimensionalization. The goodness of a solution can be measured by both correlation coefficients and resolution distributions. We present six correlation plots in Fig. 2 which measure how the predicted values be close to the labeled values, one plot per parameter. We have found that all correlation coefficients are around to one. More details could be found in the supplementary.

Apply to the $X(3872)$ and the $T_{cc}^+$ During the last decades, tens of exotic candidates have been reported [1–10]. Among them, the first and most interesting one is the $X(3872)$ which is reported by Belle Collaboration in 2003 [32]. Intensive studies have been put forward to understand its nature. For instance, the popular explanations are $D\bar{D}^* + c.c.$ hadronic molecule, compact tetra-quark and the normal charmonium with mixture of the $DD^* + c.c.$ hadronic molecule. For the detailed discussions, we would refer to Refs. [1–10]. The first two scenarios can be distinguished by the pole counting near the $DD^* + c.c.$ threshold, i.e. two poles and one pole for compact and hadronic molecules [5], respectively. These pole positions are largely related to the values of scattering length and effective range. Thus extracting these two values could help to shed light on the nature of exotic hadrons. Besides the $X(3872)$, another interesting one is $T_{cc}^+$ [17, 33] reported by LHCb in the $D^0D^0\pi^+$ channel. Since it is very close to the $D^{*+}D^0$ and $D^{*+}D^{*0}$ channels, it is viewed as a partner of the $X(3872)$ in molecular picture. In the isospin limit, i.e. neglecting the mass difference between charged and neutral charmed mesons, the $X(3872)$ and the $T_{cc}^+$ are only one-channel case, i.e. the $DD^* + c.c.$ and $DD^* + c.c.$ channel, respectively. Thus, we take them as an illustration of the applicability of our network. Although, their isospin breaking effect has several impact on physical observables [34–43], as the first step, we start from the one channel case and check the applicability.

Our network is applied to the experimental data of the $X(3872)$ [32] and the $T_{cc}^+$ [17, 33] with the three-body phase space subtracted, i.e. the $S$-wave $J/\psi\pi^+\pi^-$ and the $S$-wave $D^0\bar{D}^0\pi^+$ phase space, respectively. The obtained parameters of the two states are collected in Tab. I and Tab. II, respectively, comparing to those from fitting directly with Eq. (12). As shown in the two tables, the values of scattering length, effective range and relevant threshold from the two methods are consistent with each other within $1\sigma$ uncertainty. The resolution parameter $\sigma$ has a large deviation, which is because it is regressed individually and has larger uncertainty than those of scattering length and effective range. Especially, the absolute value of effective range of the $T_{cc}^+$ from the two methods are not as large as that in Ref. [17, 33]. The importance of this value is largely related to the nature of the $T_{cc}^+$ [15, 17–19] . As the result, extracting this parameter precisely is valuable. Our network can also extract the most relevant threshold simultaneously.

Conclusion. We train a network to analyze the experimental mass spectrums of exotic states. The 150000 data samples are generated based on Effective Range Expansion and used for training the network. As the result, the network works for near-threshold exotic states, for
which requirement most of exotic states satisfy. In addition, we apply our network for the $X(3872)$ and the $T^{c}_{cc}$ in the isospin limit. The obtained scattering length and effective range are consistent with those obtained from fit to the data directly. The advantages of our model are that all the theoretical formulas have been included in the network and the most relevant threshold can also be extracted simultaneously. This network can be used for all the one-channel near-threshold exotic states.

Acknowledgement This work is partly supported by Guangdong Major Project of Basic and Applied Basic Research No. 2020B0301030008, the National Natural Science Foundation of China with Grant No. 12035007, Guangdong Provincial funding with Grant No. 2019QN01X172. Q.W. is also supported by the NSFC and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the funds provided to the Sino-German Collaborative Research Center TRR110 “Symmetries and the Emergence of Structure in QCD” (NSFC Grant No. 12070131001, DFG Project-ID 196253076-TRR 110).
[42] M. B. Voloshin, Phys. Rev. D 76, 014007 (2007) [arXiv:0704.3029 [hep-ph]].
[43] N. A. Tornqvist, Phys. Lett. B 590, 209-215 (2004) [arXiv:hep-ph/0402237 [hep-ph]].
[44] Rene Brun and Fons Rademakers, ROOT - An Object Oriented Data Analysis Framework, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86.
[45] J. M. Clavijo, P. Glaysher, J. Jitsev and J. M. Katzy, Mach. Learn. Sci. Tech. 3, no.1, 015014 (2022) [arXiv:2005.00568 [stat.ML]].
SUPPLEMENTARY

Generation of training datasets

The training/testing datasets are generated with the Monte-Carlo technique based on the open source software ROOT [44]. The probability density function is defined as

\[
PDF(E; a, r, \text{threshold}, \sigma) = \int |T_{NR}(E)|^2 G(E' - E)dE'
\]

(11)

with

\[
T_{NR}(E) = -\frac{2\pi}{\mu} \frac{1}{1/a + (r/2)k^2 - ik},
\]

(12)

where \(\mu = \frac{m_1 m_2}{m_1 + m_2}\), and \(E\) are the reduced mass and total energy of the two-particle system. Accordingly, \(k = \sqrt{2\mu(E - m_1 - m_2)}\) is the three momentum of the scattering particle in the center-of-mass frame. \(m_1\) and \(m_2\) are the masses of the two particles.

\[
G(E) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(E - \mu)^2}{2\sigma^2}}
\]

(13)

is a Gaussian function with the mean value \(\mu\) is set to zero. \(\sigma\) is the resolution which depends on the energy resolution of measuring the invariant mass spectrum. Based on this formula, we generate 150000 line shapes for training with the parameters within the region

\[
a \in [4.93, 14.80] \text{ fm},
\]

(14)

\[
r \in [0.49, 0.99] \cup [-9.87, -0.49] \text{ fm},
\]

(15)

\[
m_1 + m_2 \in [2.8, 3.9] \text{ GeV},
\]

(16)

\[
\sigma \in [0.5, 10] \text{ MeV}.
\]

(17)

The regions of scattering length and effective range allow for both bound and virtual states [26]. The threshold region covers charmonium(-like) energy region and resolution locates usually experimental region. The four parameters are vectorized as:

\[
\text{parameters vector} = (a, r, \text{threshold}, \sigma)
\]

(18)

Within the ranges Eqs. (14),(15),(16),(17), 150000 samples of the parameters vector and corresponding histograms are uniformly generated. 45000 samples are used for testing the performance after training. These samples are indexed as:

\[
datasets = \{H_i, a_i, r_i, \text{threshold}_i, \sigma_i\}, i = 1, ..., 150000.
\]

(19)

where \(H_i\) represents a histogram hosting 100 paired values, i.e., the mass spectrum. Fig. 3 illustrates uniform distributions of the parameter \(\sigma\) and threshold. Fig. 4 illustrates 2D histogram for the parameter \(a\) and \(r\) (the left column). Given a specific value of parameter vector, the example mass spectrums are illustrated in the right column of Fig. 4.
FIG. 4. The left panels show 2D histogram for the parameter $a$ and $r$ at generation level in cases of a bound state (a), a resonance (b) and a virtual state (c). The right panels show 200 data points illustrated in histograms for a bound state (a), a resonance (b), and a virtual state (c), respectively, for a specific value of parameters vector.
Training

A multi-layer-perception [27] based ResNet [28] is implemented with the open source PyTorch[31] framework as illustrated in Fig. 1. This model consists of three ResBlocks, one input layer accepting our histogram (200-dimensional vector) and one output layer connecting to the label values. The ResBlock consists of two fully-connected layers accompanied by dropout layers. The input layer of the ResBlock has a shortcut connection to the output layer. In this way, solving this model is of high efficiency.

At the beginning of training, the model needs to be initialized. The weights of neurons are randomly initialized with a normal distribution, and the biases of neurons are set to zero. The threshold value of dropout layers are set to 0.3. The label values for the parameters \(a\) and \(r\) are applied with normalization and nondimensionalization as below:

\[
a_{\text{norm}} = \frac{a_{\text{generation}}}{a_{\text{max}}} \tag{20}
\]

\[
r_{\text{norm}} = \frac{r_{\text{generation}} - r_{\text{min}}}{r_{\text{max}} - r_{\text{min}}} \tag{21}
\]

These two parameters are simultaneously regressed because they are correlated in case of a resonance state. While the threshold and \(\sigma\) are individually regressed since they are independent.

The next, we choose the Adam [30] optimizer to solve above model, and the MSELoss function (the mean squared error loss) to measure the Euclidean distance between the prediction values and the label values. The Adam [30] optimizer is one of the most widely used optimizers which combines the Momentum algorithm and the RMSProp algorithm[45]. It not only fasten the convergence but also reduce the fluctuation of the loss function. A reasonable solution could be obtained around 1000 training epochs using an initial learning rate value of 0.001, which is automatically and dynamically adjusted during the training cycle. As illustrated in Fig. 5, we have found that the MSELoss function value converges rapidly in two hundred iterations for the regression of which parameter.

FIG. 5. The MSELoss function converges as training epoches increase.
Evaluation

We have presented the correlation plots in Fig. 2, and have found that the correlation coefficients are around to one. We further extract the difference distributions between the predicated values and the label values as shown in Fig. 6 (the left column), in which plots (a1-d1) are for the parameters a, r, threshold, and $\sigma$. These distributions are obtained by testing 45000 samples. The mean measures the prediction values deviation from the label values, and the RMS measures the intrinsic uncertainties of this method. Relevant numbers are summarized in the Table III. For a straightforward comparison, we extract the parameters from directly fitting to our testing samples. 15000 samples were tried, among which only 11337 fitting passed with an acceptable $\chi^2$ less than 100. For these successful fitting, the parameters’ values are filled in four distributions as shown in Fig. 6 (the right column). We have found that biases of the deep learning could be neglected for all parameters, and intrinsic uncertainties could be neglected for the parameters a, threshold and $\sigma$.

| methods   | Deep learning | Fitting    |
|-----------|----------------|------------|
| parameters↓ | bias uncertainty | bias uncertainty |
| a (fm)    | -0.010         | 1.040     |
| r (fm)    | -0.033         | 0.268     |
| threshold (MeV) | 0.75         | 0.52      |
| $\sigma$ (MeV) | -0.0001       | 0.06      |

Table III. The biases and errors information of models
FIG. 6. Distributions of the difference between prediction values and label values for each parameter. Distributions in the left column are obtained with the deep learning while those in the right column with the fitting method.