Wuhan University
Journal of Natural Sciences

2022, Vol.27 No.2, 161-168

Article ID 1007-1202(2022)02-0161-08
DOI https://doi.org/10.1051/wujns/2022272161

Using Deep Learning Algorithms to Improve Energy Resolution in the Semileptonic Decays

☐ **WANG Yang, CAI Hao†, SUN Liang†**

School of Physics and Technology, Wuhan University, Wuhan 430072, Hubei, China

Abstract: The neutrino closure method can be used to obtain the decay kinematics with one missing final state particle (v) in semileptonic decays. Its solution should give the square of the invariant mass of the lv system (q^2) and momentum (P) of the decayed mother particle in semileptonic decay process. However, the resolution obtained by solving two-solution problems with existing algorithms is limited. We propose a new method based on deep learning to improve the resolution of the two key physical quantities when processing Large Hadron Collider beauty (LHCb) experimental data. Resolution of q^2 (P) can be improved evenly 1.7% (8.2%) by regression algorithm and 2.7% (9.6%) by classification algorithm compared to linear regression algorithm. The resolution improvements using the new method will benefit the studies on semileptonic decays in hardon collider experiments. Moreover, the new method can be applied to other decays with a missing particle in the final state.

Key words: Large Hadron Collider beauty (LHCb) experiment; semileptonic decay; improving resolution; deep neural network
CLC number: O572.21+3

Received date: 2022-02-10
Foundation item: Supported by the National Natural Science Foundation of China (11735010, U1932108, U2032102, 12061131006)
Biography: WANG Yang, male, Master candidate, research direction: particle physics experiment. E-mail: wangya@whu.edu.cn
† To whom correspondence should be addressed.hcai@whu.edu.cn; sunl@whu.edu.cn

0 Introduction

Large Hadron Collider beauty (LHCb) experiment[1] has the advantages of high energy and large data statistics, and it also provides an excellent opportunity to study semileptonic decay[2]. However, great challenges are still required to unravel. Determining the square of the invariant mass (q^2) of the lv system (lepton and neutrino system) in the hadron collider experiment is a challenging and important task. The form factors in decay are constituted by q^2, improving the precision of form factor via the improvement of resolution of q^2. The Cabibbo-Kobayashi-Maskawa (CKM) matrix elements can be extracted in a more precise way based on the resolution improvement of q^2[3-5].

The q^2 can be researched via momentum of particle in semileptonic decay process. But the neutrino cannot be detected because of strong penetration, and then momentum of the parent particle does not include the component of neutrino in semileptonic decay. Neutrino closure method[6] can be used to reconstruct the momentum of decayed parent particle in semileptonic decay processes, which is based on topological information and mass hypothesis. It can be applied in the research of q^2 in order to better solve the impact of missing neutrino on decayed parent momentum. However, this method has a limitation of the two-solution problem in quadratic equation. The original method is selected randomly by random seed, but the resolution of q^2 is not satisfactory. Recently, a novel method[7] based on linear regression algorithm (LR) has been proposed to solve the two-solution

problem. It makes sense to find ways to enhance further the resolution improvement of q^2 and the momentum of the decayed parent particle.

This paper proposes a new idea based on deep neural network to solve the two-solution problem in neutrino closure method. The error of form factor and CKM elements in semileptonic decay are expected to be further reduced by our new idea.

1 Simulation of Semileptonic Decay

In this paper, RapidSim[8] is used to generate the sample events of different decay processes in the environment of LHCb[9] , which request acceptance is "All-DownStream". Pythia 8.2[10] is used to produce *pp* collisions sample events at the center of mass energy of 13 TeV. The decay mode of hadrons is described by EvtGen[11] and the radiation of the final state particle described by PHOTOS[12] .

Different semileptonic decay processes in LHCb experiment we selected in this research are $B_s^0 \to K^-\mu^+\upsilon_\mu$ [13], $B_s^0 \to D^{*-}\mu^+\upsilon_\mu$ ($D^{*-} \to D^0\pi^-$, $D^0 \to K^-\pi^+$)[14], $\Lambda_b^0 \to p^+\mu^-\overline{\upsilon}_\mu$ [15] and $D^+ \to K^-\pi^+\mu^+\upsilon_\mu$. We define the momentum of particle as P, the transverse momentum relative to the Z-axis as P_T and the component of momentum along the Z-axis as P_z by $P_T = P\cdot\sin\theta$ in this simulation. The pseudorapidity of a particle is defined as $\eta = -\ln(\tan(\theta/2))$. In order to make the generated sample meet the acceptance requirement of the LHCb[16], the pure sample events are selected by pseudorapidity (η), transverse momentum (P_T) and momentum (P) of particles for different decay processes. The signal instance requests $2<\eta<5$ for B_s^0 and Λ_b^0 , and for D^+ requests $1.7<\eta<5.2$ to be met in LHCb detector[17]. The distribution of η of B_s^0 , Λ_b^0 and D^+ are shown in Fig. 1. The selection conditions for different decay processes[18,19] are shown in Table 1. Meanwhile, the distribution of P and P_T are shown in Fig. 2 and Fig. 3. The natural unit c(light speed) =1 is used in this paper. The different decay process will use other bracketed number expressed in Table 1. "(1)" is used to represent $B_s^0 \to K^-\mu^+\upsilon_\mu$ decay process, "(2)" is used to represent $\Lambda_b^0 \to p^+\mu^-\overline{\upsilon}_\mu$ decay process, "(3)" is used to represent $D^+ \to K^-\pi^+\mu^+\upsilon_\mu$ and, "(4)" is used to represent $B_s^0 \to D^{*-}\mu^+\upsilon_\mu$ decay process.

2 Variables Used to Correct Momentum of Particles

Before reconstructing the q^2, the momentum of parent particle must be reconstructed. The final state particles generated by parent particle decay process, the momentum of final particles, have a direct relationship with momen

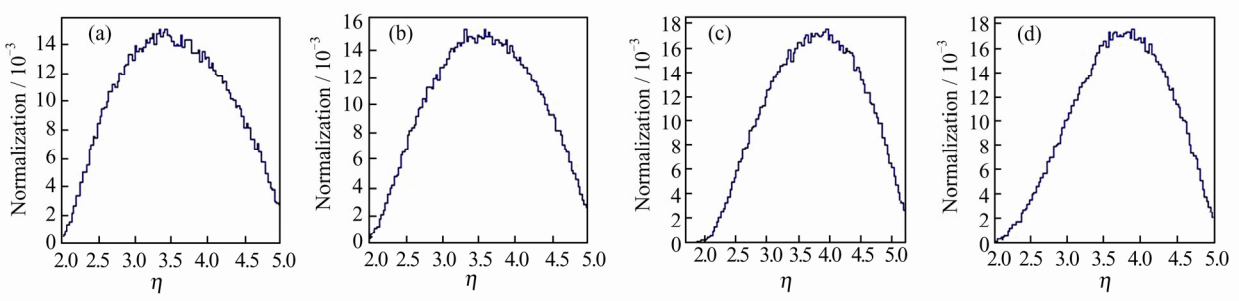

Fig. 1 The distribution of pseudorapidity (η) of decayed particles

(a) $B_s^0 \to K^-\mu^+\upsilon_\mu$, (b) $\Lambda_b^0 \to p^+\mu^-\overline{\upsilon}_\mu$, (c) $D^+ \to K^-\pi^+\mu^+\upsilon_\mu$ and (d) $B_s^0 \to D^{*-}\mu^+\upsilon_\mu$

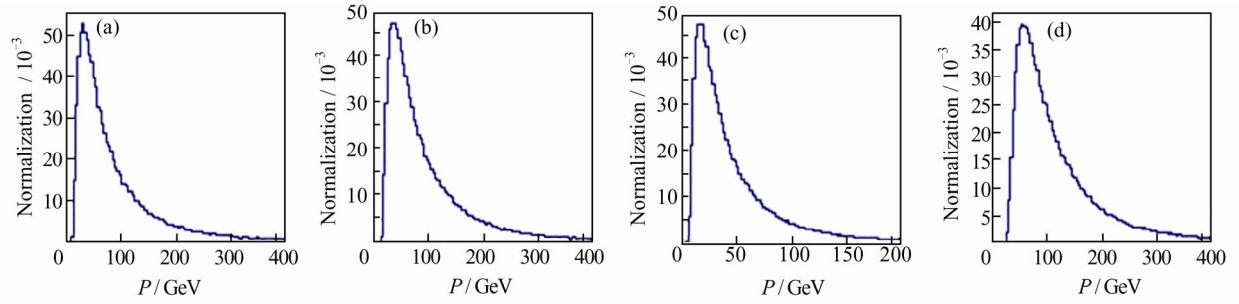

Fig. 2 The distribution of momentum of decayed particles for total decay processes in sample events

(a) $B_s^0 \to K^-\mu^+\upsilon_\mu$, (b) $\Lambda_b^0 \to p^+\mu^-\overline{\upsilon}_\mu$, (c) $D^+ \to K^-\pi^+\mu^+\upsilon_\mu$ and (d) $B_s^0 \to D^{*-}\mu^+\upsilon_\mu$; The natural unit c (light speed)=1

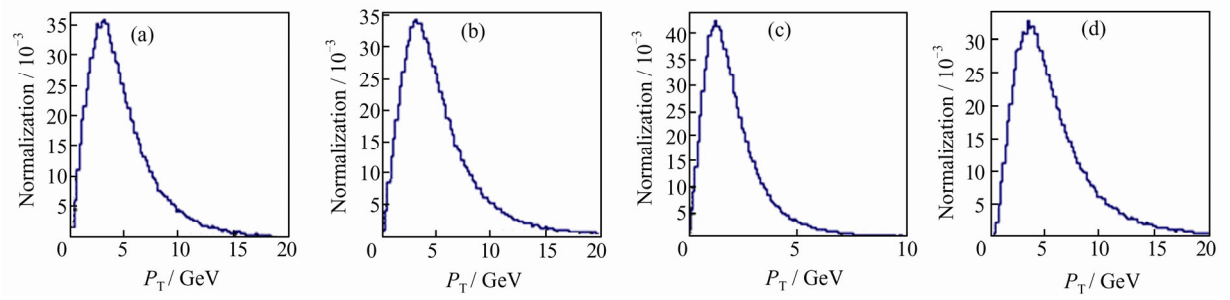

Fig. 3 The distribution of transverse momentum of decayed particles for total decay processes in sample events
(a) $B_s^0 \to K^-\mu^+\upsilon_\mu$, (b) $\Lambda_b^0 \to p^+\mu^-\overline{\upsilon}_\mu$, (c) $D^+ \to K^-\pi^+\mu^+\upsilon_\mu$ and (d) $B_s^0 \to D^{*-}\mu^+\upsilon_\mu$; The natural unit c (light speed)=1

Table 1 Selection condition for total decay processes

Decay process	Particle	η	P_T / GeV	P / GeV
(1)	B_s^0	$2<\eta<5$	—	—
	K^-	$1.9<\eta<4.9$	$P_T>0.5$	$P>5$
	μ^+	$1.9<\eta<4.9$	$P_T>1$	$P>5$
(2)	Λ_b^0	$2<\eta<5$	—	—
	p^+	$1.9<\eta<4.9$	$P_T>0.5$	$P>9.3$
	μ^-	$1.9<\eta<4.9$	$P_T>1$	$P>5$
(3)	D^+	$1.7<\eta<5.2$	—	—
	π^+	$1.7<\eta<5.2$	$0.3<P_T<40$	$P>1$
	K^-	$1.7<\eta<5.2$	$0.8<P_T<40$	$P>3$
	μ^+	$1.7<\eta<5.2$	$0.5<P_T<40$	—
(4)	B_s^0	$2<\eta<5$	—	—
	D^{*-}	$2<\eta<5$	$P_T>0.25$	$P>5$
	μ^+	$2<\eta<5$	$P_T>0.25$	$P>5$

tum of the parent particle because of momentum conservation. Meanwhile three-momentum and mass information of the final state particles are important for reconstructing the 4-D Lorentz vector of parent particle for every decay process. So the simulated three-momentum (P_x, P_y, P_z) of final state particles will be chosen as a part of the input features.

As mentioned in the research of $B_s^0 \to K^-\mu^+\upsilon_\mu$ [13], the corrected momentum of the B_s^0 can be predicted by linear regression (LR) using the variables related to the flight vector. Flight vector is crucial because it is used to model the resolution of the momentum and calculate the corrected mass. Another part of input features we need to consider in this study is the line between the primary vertex and the decay vertex of the decayed particles. Due to the characteristics of the nonlinear model, we directly select the coordinate of the primary vertex and the end vertex on X, Y, Z as a part of the input features. Meanwhile, the angle between P and P_z of particles in the spherical coordinate system are chosen in order to reduce the smear effect of the primary vertex and end vertex. The same strategy will be used for the other three decay processes.

The impact parameters of final state particles are chosen to be the input variables because the impact parameter is critical to the multiplicity and average transverse momentum of the particles produced by the collision.

3 Introduction of Deep Neural Network Algorithm

3.1 Deep Neural Network Regression Algorithm

Fully connected regression deep neural network is chosen to predict momentum. The total variables we mentioned in Section 2 are regarded as the input feature

of fully connected deep neural network regression model (DNNR). The structure of DNNR constructed by us consists of four layers with the PReLU[20] activation, and its optimizer is Adam[21] in order to predict the momentum of decayed parent particles more accurately, the model is based on Keras[22] and Tensorflow package[23]. The details of regression algorithm are shown in Algorithm 1.

Algorithm 1 Deep neural network regression algorithm used for different decay processes.

Input: the original feature variables dataset I^0, size is $n_v \times n_e$, n_v is the number of input variables, n_e is the number of events

Output: the regression results (predicted momentum) for testing dataset y^t.

1) **for each group** input dataset I_i, construct it by the value of input variables for one event, exist n_e groups

$$I^0 = \{I_i\}_{i=1,2,\cdots,n_e}$$

$$I_i = \{V_j\}_{j=1,2,\cdots,n_v}$$

2) **for each group** target dataset, construct by true momentum of decayed parent particle,

$$\{y_k^{\text{true}}\}_{k=1,2,\cdots,n_e}$$

3) **for each** I_i in I^0

(a) input the model of 4 layer fully connected neural regression network, and PReLU activation is used in the middle of different layers;

(b) train the model by Adam optimization;

(c) obtain the regression result of training dataset $\{y_k^{\text{train}}\}_{k=1,2,\cdots,n_e}$ and the weight parameters of the model.

4. calculate the RMS between $\{y_k^{\text{train}}\}_{k=1,2,\cdots,n_e}$ and $\{y_k^{\text{true}}\}_{k=1,2,\cdots,n_e}$ for deciding the performance of the model;

5. for each I_m in test dataset I^t, input it to the trained model;

6. return regression result (predicted momentum) of I^t, y^t.

Testing samples generated by using the same option with the training samples, training samples are used to train machine learning model and testing samples are used to test model and do physical calculation. The Scikit-learn package[24] is used to split sample randomly. The uproot4 package, one part of the HepML[25] package, is used to import data to the machine learning models. The Numpy package[26] will be used to transform data to the array when validating the model. Meanwhile, calculation of correlation coefficients is supported by the PANDAS package[27]. The training process is based on NVIDIA GeForce RTX 2080ti for speed. Figure 4 shows the data stream sketch map of the DNNR.

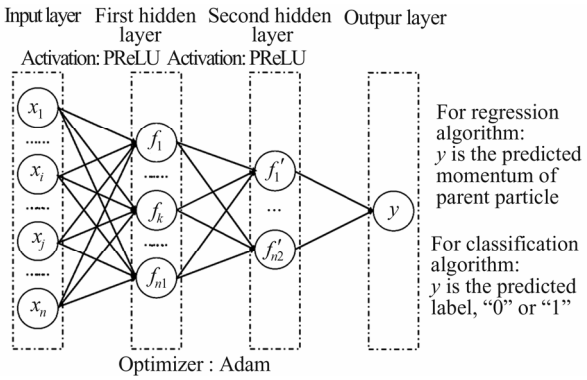

Fig. 4 The flow chart of deep neural network regression algorithm and classification algorithm

The correlation coefficients (ρ) of P_{pred} (momentum predicted by DNNR (P_{DNN}) and LR (P_{Linear})) versus the true momentum (P_{true}) of decayed particle for different decay processes are shown in Table 2. The more information the model gets, the closer the distribution of P_{pred} is to the distribution of P_{true}. So the correlation between P_{DNN} and P_{true} is higher than that between P_{Linear} and P_{true}. Although nonlinear model can better describe the distribution of target compared with the linear model, the over fitting of the model is worthy of attention at the same time.

Table 2 The correlation coefficient (ρ) of P_{pred} (P_{DNN}, P_{Linear}) *vs.* P_{true} for total decay processes

Decay process	P_{DNN} *vs.* P_{true}	P_{Linear} *vs.* P_{true}
$B_s^0 \to K^- \mu^+ \upsilon_\mu$	0.907	0.705
$\Lambda_b^0 \to p^+ \mu^- \overline{\upsilon}_\mu$	0.917	0.692
$D^+ \to K^- \pi^+ \mu^+ \upsilon_\mu$	0.931	0.761
$B_s^0 \to D^{*-} \mu^+ \upsilon_\mu$	0.933	0.707

According to Fig. 5, the distribution, (P_{pred}–P_{true})/P_{true}, which is predicted by DNNR, is better than that predicted by LR on different decay processes. The RMS of using DNNR is 35.26% better than using LR for $B_s^0 \to K^- \mu^+ \upsilon_\mu$, for $\Lambda_b^0 \to p^+ \mu^- \overline{\upsilon}_\mu$ is 36.12%, for $D^+ \to K^- \pi^+ \mu^+ \upsilon_\mu$ is 39.23% and for $B_s^0 \to D^{*-} \mu^+ \upsilon_\mu$

is 33.54% from Table 3. Therefore, it can be learned that the distribution of momentum predicted by the DNNR model is closer to the distribution of true momentum than LR. This also means that the DNNR will perform better than LR on selecting a solution closer to the true value in neutrino closure method.

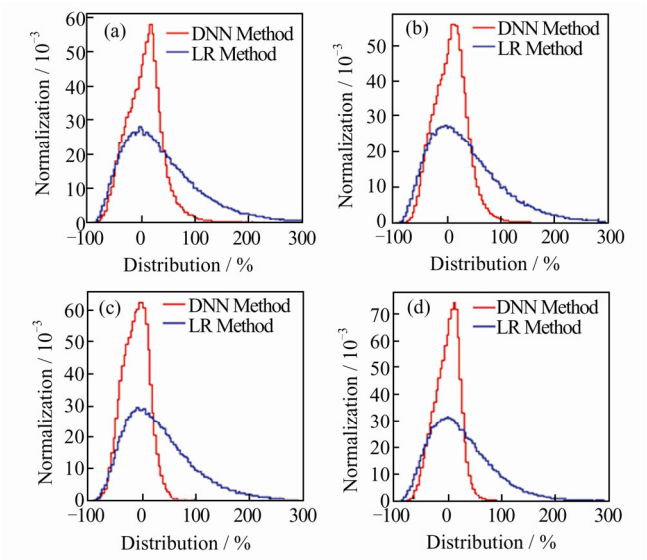

Fig. 5 Comparison of distribution $(P_{\text{pred}}-P_{\text{true}})/P_{\text{true}}$ for LR and DNNR

(a) $B_s^0 \to K^-\mu^+\upsilon_\mu$, (b) $\Lambda_b^0 \to p^+\mu^-\overline{\upsilon}_\mu$, (c) $D^+ \to K^-\pi^+\mu^+\upsilon_\mu$ and (d) $B_s^0 \to D^{*-}\mu^+\upsilon_\mu$

Table 3 RMS of $(P_{\text{pred}}-P_{\text{true}})/P_{\text{true}}$ of total decay processes

%

Decay process	RMS of using LR	RMS of using DNNR	ΔRMS
$B_s^0 \to K^-\mu^+\upsilon_\mu$	34.01	69.27	35.26
$\Lambda_b^0 \to p^+\mu^-\overline{\upsilon}_\mu$	30.60	66.72	36.12
$D^+ \to K^-\pi^+\mu^+\upsilon_\mu$	24.63	64.86	39.23
$B_s^0 \to D^{*-}\mu^+\upsilon_\mu$	23.30	56.84	33.54

3.2 Deep Neural Network Classification Algorithm

In addition to using the deep neural network regression algorithm for the momentum correction, we also propose another new method based on fully connected deep neural network classification algorithm (DNNC), by regarding the situation of two-solutions in the neutrino closure method as a mathematically separate problem. The two solutions can be regarded as two classes, and every solution will be given a label which is "0" or "1". We make a classification model to select the better solution according to give the label defined. The construction of the model uses the MLP Classifier model of Scikit-learn package based on CPU, and the structure is the same as DNNR that consists of one input layer, two hidden layers, and one output layer. The activation is default option-ReLU[28] and optimizer is also Adam (same as DNNR). The details of the algorithm used by us are shown in Algorithm 2.

Algorithm 2 Deep neural network classification algorithm used for different decay processes.

Input: the original feature variables dataset I^{o}, size is $n_{\text{v}} \times n_{\text{e}}$, n_{v} is the number of input variables, n_{e} is the number of events

Output: the classification results for testing dataset R^t.

1. **for each group** input dataset I_i, construct it by the value of input variables for one event, exist n_{e} groups.

$$I^{\text{o}} = \{I_i\}_{i=1,2,\cdots,n_{\text{e}}}$$

$$I_i = \{V_j\}_{j=1,2,\cdots,n_{\text{v}}}$$

2. **for each group target** dataset, use true momentum to select a better result from two solutions for every event, two solutions are $x = \dfrac{-b \pm \sqrt{b^2-4ac}}{2a}$; if the better result satisfy "+" ("−"), the label is "1" ("0"), target dataset are constructed by the label for total events

$$R^{\text{true}} = \{R_i\}_{i=1,2,\cdots,n_{\text{e}}}$$

3. **for each** I_i in I^{o}

   (a) input the model of 4 layer fully connected neural network, ReLU activation is used in the middle of different layers;

   (b) train the model by Adam optimization;

   (c) obtain the classification result (predicted label) of training dataset $\{R_k^{\text{train}}\}_{k=1,2,\cdots,n_{\text{e}}}$ and the weight parameters of the model;

4. validate the accuracy between $\{R_k^{\text{train}}\}_{k=1,2,\cdots,n_{\text{e}}}$ and $\{R_i\}_{i=1,2,\cdots,n_{\text{e}}}$ for deciding the robustness of the model;

5. for each I_m (each group) in test dataset I^t, input it to the trained model;

6. return classification result of I^t, R^t.

Compared with the regression algorithm, the classification algorithm leaves out the middle process of predicting the inferred momentum given by model, and gives the final corrected momentum result directly. The fully connected deep neural network regression algorithm has two parts of errors, the predicting momentum process and selection process. Fully connected deep neural network classification only has one part of error because of the more direct process. The structure of DNNC model is shown in Fig. 4.

4 Physics Application to the Study of Semileptonic Decay Processes

The target of using the DNNR or DNNC to predict the momentum of the decayed particles is to select a better one from the two solutions in the neutrino closure method[6] on different decay processes. The sum momentum vector of the reconstructed system by detector is defined as the visible momentum, which is denoted $\vec{P}_{\text{vis}}$. The momentum of the invisible neutrinos from a semileptonic decay is denoted as $\vec{P}_{\text{miss}}$. The visible momentum is decomposed into two parts, the perpendicular component ($P_{\text{vis}}^{\perp}$) and parallel component ($P_{\text{vis}}^{\parallel}$)[7]:

$$P_{\text{vis}}^{\perp} = \left| \vec{P}_{\text{vis}} \times \frac{\vec{F}}{|\vec{F}|} \right| \tag{1}$$

$$P_{\text{vis}}^{\parallel} = \vec{P}_{\text{vis}} \cdot \frac{\vec{F}}{|\vec{F}|} \tag{2}$$

In the specified coordinate system, the perpendicular component of the missing momentum is required to be equal to the perpendicular component of the visible momentum, $P_{\text{miss}}^{\perp} = P_{\text{vis}}^{\perp}$. Assuming the mass of decayed parent particle is *m*, and there is a single massless unreconstructed particle (neutrino) in the final state, $P_{\text{miss}}^{\parallel}$ can be obtained by the solutions of the quadratic equation. The α and β of the quadratic equation are defined as:

$$\alpha = \frac{P_{\text{vis}}^{\parallel}(m^2 - M_{\text{vis}}^2 - 2(P_{\text{vis}}^{\perp})^2)}{2((P_{\text{vis}}^{\parallel})^2 - E_{\text{vis}}^2)} \tag{3}$$

$$\beta = \frac{E_{\text{vis}}^2(m^2 - M_{\text{vis}}^2 - 2(P_{\text{vis}}^{\perp})^2)}{4((P_{\text{vis}}^{\parallel})^2 - E_{\text{vis}}^2)^2} + \frac{(E_{\text{vis}}P_{\text{vis}}^{\perp})^2}{(P_{\text{vis}}^{\parallel})^2 - E_{\text{vis}}^2} \tag{4}$$

E_{vis} is the visible energy and M_{vis} is the visible mass. $P_{\text{miss}}^{\parallel}$ is defined as:

$$P_{\text{miss}}^{\parallel} = -\alpha \pm \sqrt{\beta} \tag{5}$$

This yields two solutions for the momentum of neutrino closure method:

$$P_{+} = P_{\text{vis}}^{\parallel} - \alpha + \sqrt{\beta} \tag{6}$$

$$P_{-} = P_{\text{vis}}^{\parallel} - \alpha - \sqrt{\beta} \tag{7}$$

The distribution of $(P_{-} - P_{\text{true}})/P_{\text{true}}$ *vs.* $(P_{+} - P_{\text{true}})/P_{\text{true}}$ of different decay processes is shown in Fig. 6. A horizontal band can be seen clearly for P_{+} and the same as for P_{-} in the vertical direction. Although the effect of the vertex smearing is visible, the two bands are nevertheless well separated, and the feature between P_{+} and P_{-} is apparent.

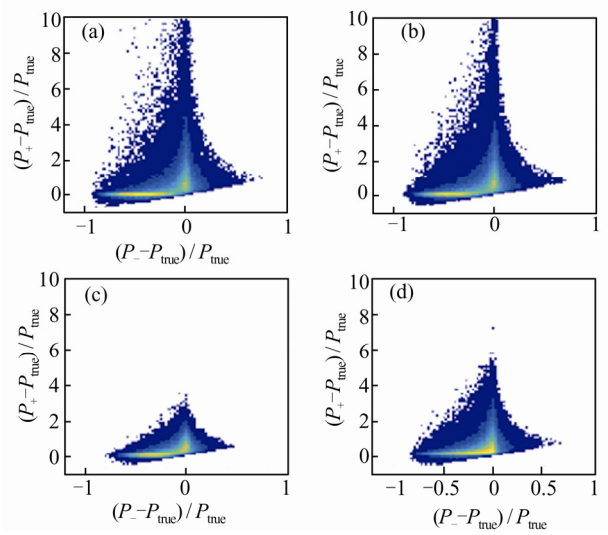

Fig. 6 The distribution of $(P_- - P_{\text{true}})/P_{\text{true}}$ *vs.* $(P_+ - P_{\text{true}})/P_{\text{true}}$
(a) $B_s^0 \to K^- \mu^+ \upsilon_\mu$, (b) $\Lambda_b^0 \to p^+ \mu^- \overline{\upsilon}_\mu$, (c) $D^+ \to K^- \pi^+ \mu^+ \upsilon_\mu$ and (d) $B_s^0 \to D^{*-} \mu^+ \upsilon_\mu$

The information used in DNNR and DNNC is more than that used in LR because of the restriction of physical formula in LR. Therefore, the performance of using DNNR and DNNC in selecting a better solution for different decay processes in neutrino closure method is better. The solution selected by true momentum of decayed particles represents the upper threshold. The closer the result obtained with the machine learning model to the upper threshlod, the better the performance of the machine learning model.

Now we can compare the performance of our new idea on q^2 and momentum for the different decay processes. According to Fig.7, the distribution using DNNR and DNNC methods is closer to the upper threshold compared with using LR for total decay processes, meaning the method based on deep neural network will perform better in reconstruction of invariant mass q^2 of the *lv* system for the semileptonic decay processes used in this paper. The resolution improvement of decayed

parent particles for different decay processes are also shown in Table 4. According to Table 5, the resolution of q^2 increases by 2.04% and 2.25% by using DNNR and DNNC than that by using LR for $B_s^0 \to K^-\mu^+\upsilon_\mu$, 2.53% and 3.84% for using DNNR and DNNC than using LR for $\Lambda_b^0 \to p^+\mu^-\overline{\upsilon}_\mu$, 0.83% and 2.67% by using DNNC than that by using LR for $D^+ \to K^-\pi^+\mu^+\upsilon_\mu$ and 1.53% and 2.02% by using DNNR and DNNC for $B_s^0 \to D^{*-}\mu^+\upsilon_\mu$. The resolution improvement of momentum and q^2 is obvious, which means the performance of deep neural network idea with either regression algorithm or classification algorithm in solving two-solution problem of neutrino closure method is better than linear regression algorithm.

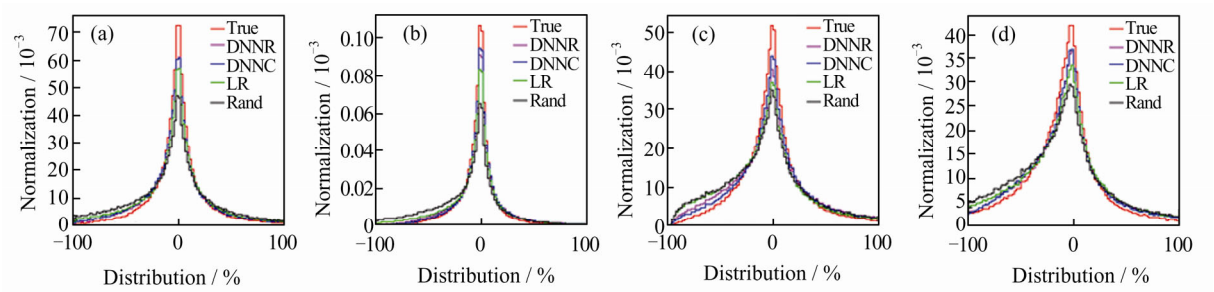

Fig. 7 Comparison of distribution $(q^2_{\rm corr} - q^2_{\rm true})/q^2_{\rm true}$

(a) $B_s^0 \to K^-\mu^+\upsilon_\mu$, (b) $\Lambda_b^0 \to p^+\mu^-\overline{\upsilon}_\mu$, (c) $D^+ \to K^-\pi^+\mu^+\upsilon_\mu$ and (d) $B_s^0 \to D^{*-}\mu^+\upsilon_\mu$. True: the q^2 obtained by true momentum of decayed parent particle to select the better solution, this result is upper threshold in this research; Rand: the q^2 obtained by random strategy to select better solution; DNNC: q^2 obtained by trained DNNC model to select the better solution; DNNR (LR): q^2 obtained by the predicted momentum ($P_{\rm pred}$) predicted by DNNR (LR) to select the better solution

Table 4 The RMS of $(P_{\rm corr} - P_{\rm true})/P_{\rm true}$ for different decay processes

%

Decay process	DNNR	DNNC	LR	DNNR improvement	DNNC improvement
$B_s^0 \to K^-\mu^+\upsilon_\mu$	30.95	30.26	37.25	6.30	6.99
$\Lambda_b^0 \to p^+\mu^-\overline{\upsilon}_\mu$	29.59	28.40	37.33	7.74	8.93
$D^+ \to K^-\pi^+\mu^+\upsilon_\mu$	23.31	25.75	32.51	9.20	6.76
$B_s^0 \to D^{*-}\mu^+\upsilon_\mu$	25.24	19.10	34.91	9.67	15.81
Mean	—	—	—	8.23	9.6

Table 5 The RMS of $(q^2_{\rm corr} - q^2_{\rm true})/q^2_{\rm true}$ for different decay processes

%

Decay process	DNNR	DNNC	LR	DNNR improvement	DNNC improvement
$B_s^0 \to K^-\mu^+\upsilon_\mu$	32.39	32.18	34.43	2.04	2.25
$\Lambda_b^0 \to p^+\mu^-\overline{\upsilon}_\mu$	26.65	25.34	29.18	2.53	3.84
$D^+ \to K^-\pi^+\mu^+\upsilon_\mu$	37.41	35.57	38.24	0.83	2.67
$B_s^0 \to D^{*-}\mu^+\upsilon_\mu$	38.04	37.55	39.57	1.53	2.02
Mean	—	—	—	1.73	2.69

5 Conclusion

Compared with using the linear regression method, the distributions of the solution selected by the methods of deep neural network regression and classification move to the upper threshold. Resolution of q^2 (P) can be improved evenly by 1.7% (8.2%) using deep neural network regression algorithm and 2.7% (9.6%) by deep neural network classification algorithm compared with linear regression algorithm. Accompanied by improving the resolution of momentum, the effect caused by missing neutrino can be reduced. The momentum reconstructed by the neutrino closure method is close to the actual situation, which will supply a good chance to better study the kinematics of semileptonic decay process. The improvement of resolution of q^2 will also lead to better constraints on the improvement of precision of forms factors, providing a chance to calculate the branch fractions more accurately for undiscovered semileptonic decay processes and improve the knowledge of CKM matrix elements.

References

[1] Alves Jr A A, Andrade Filho L M, Barbosa A F, *et al*. The LHCb detector at the LHC [J]. *Journal of Instrumentation*, 2008, **3**(8): S08005.

[2] Dominik M. *Study of Semileptonic D^0 Decays for a Measurement of Charm Mixing at LHCb* [D]. Heidelberg: University of Heidelberg, 2014.

[3] Ablikim M, Achasov M N, Adlarson P, *et al*. Future physics programme of BESIII [J]. *Chinese Physics C*, 2020, **44**(4): 040001.

[4] Na H, Davies C T, Follana E, *et al*. $D \to K,l\upsilon$ semileptonic decay scalar form factor and $|V_{cs}|$ from lattice QCD [J]. *Physical Review-Section D-Particles and Fields*, 2011, **84** (11): 114506.

[5] Shen Y L, Wei Y B. $B \to P,V$ form factors with the B-meson light-cone sum rules [J]. *Advances in High Energy Physics*, 2022, **2022**: 2755821.

[6] Dambach S, Langenegger U, Starodumov A. Neutrino reconstruction with topological information [J]. *Nuclear Instruments and Methods in Physics Research Section A*: *Accelerators*, *Spectrometers*, *Detectors and Associated Equipment*, 2006, **569**(3): 824-828.

[7] Ciezarek G, Lupato A, Rotondo M, *et al*. Reconstruction of semileptonically decaying beauty hadrons produced in high energy pp collisions [J]. *Journal of High Energy Physics*, 2017, **2017**(2): 21.

[8] Cowan G A, Craik D C, Needham M D. RapidSim: An application for the fast simulation of heavy-quark hadron decays [J]. *Computer Physics Communications*, 2017, **214**: 239-246.

[9] Aaij R, Albrecht J, Alessio F, *et al*. The LHCb trigger and its performance in 2011 [J]. *Journal of Instrumentation*, 2013, **8**(4): P04022.

[10] Sjöstrand T, Ask S, Christiansen J R, *et al*. An introduction to PYTHIA 8.2 [J]. *Computer Physics Communications*, 2015, **191**: 159-177.

[11] Lange D J. The EvtGen particle decay simulation package [J]. *Nuclear Instruments and Methods in Physics Research Section A*: *Accelerators*, *Spectrometers*, *Detectors and Associated Equipment*, 2001, **462**(1-2): 152-155.

[12] Golonka P, Kersevan B, Pierzchała T, *et al*. The tauola-photos-F environment for the TAUOLA and PHOTOS packages, release II [J]. *Computer Physics Communications*, 2006, **174**(10): 818-835.

[13] Aaij R, Beteta C A, Ackernley T, *et al*. First observation of the decay $B_s^0 \to K^-\mu^+\upsilon_\mu$ and a measurement of $|V_{ub}|/|V_{cb}|$ [J]. *Physical Review Letters*, 2021, **126**(8): 081804.

[14] Flynn J, Hill R, Jüttner A, *et al*. Semileptonic $B \to \pi l\upsilon$, $B \to Dl\upsilon$, $B_s \to Kl\upsilon$, and $B_s \to D_s l\upsilon$ decays [J]. *Proceedings of Science*, 2020, **363**: 184.

[15] Aaij R, Adeva B, Adinolfi M, *et al*. Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays [J]. *Nature Phys*, 2015, **11**: 743-747.

[16] Aaij R, Beteta C A, Adeva B, *et al*. Measurement of b hadron fractions in 13 TeV pp collisions [J]. *Physical Review D*, 2019, **100**(3): 031102.

[17] Antunes Nobrega R, Franca Barbos A, Bediaga I, *et al*. LHCb reoptimized detector design and performance: Technical design report [EB/OL]. [2021-12-10]. *https://hal. archives-ouvertes.fr/in2p3*-00025912.

[18] Detmold W, Lehner C, Meinel S. $\Lambda_b \to pl^-\overline{\upsilon}_l$ and form factors from lattice QCD with relativistic heavy quarks [J]. *Physical Review D*, 2015, **92**(3): 034503.

[19] Aaij R, Beteta C A, Ackernley T, *et al*. Measurement of $|V_{cb}|$ with $B_s^0 \to D_s^{(*)-}\mu^+\upsilon_\mu$ decays [J]. *Physical Review D*, 2020, **101**(7): 072004.

[20] He K M, Zhang X Y, Ren S Q, *et al*. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification[C]//2015 *IEEE International Conference on Computer Vision*. New York: IEEE, 2015: 1026-1034.

[21] Chang Z H, Zhang Y, Chen W B. Effective Adam-optimized LSTM neural network for electricity price forecasting [C]// 2018 *IEEE* 9*th International Conference on Software Engineering and Service Science*. New York: IEEE, 2018: 245- 248.

[22] Ketkar N. *Deep Learning with Python* [M]. Berkeley: Apress, 2017: 97-111.

[23] Abadi M, Barham P, Chen J, *et al*. TensorFlow: A system for large-scale machine learning [C]//12*th USENIX Symposium on Operating Systems Design and Implementation* (*OSDI* 16). New York: ACM, 2016: 265-283.

[24] Pedregosa F, Varoquaux G, Gramfort A, *et al*. Scikit-learn: machine learning in python [J]. *The Journal of Machine Learning Research*, 2011, **12**: 2825-2830.

[25] Belov S, Dudko L, Kekelidze D, *et al*. HepML, an XML-based format for describing simulated data in high energy physics [J]. *Computer Physics Communications*, 2010, **181** (10): 1758-1768.

[26] Oliphant T E. *A Guide to NumPy* [M]. New York: Trelgol Publishing, 2006.

[27] Snider L A, Swedo S E. PANDAS: Current status and directions for research [J]. *Molecular Psychiatry*, 2004, **9**(10): 900-907.

[28] Hanin B. Universal function approximation by deep neural nets with bounded width and Relu activations [J]. *Mathematics*, 2019, **7**(10): 992.

□