Changing Parameters in the Deep Neural Network based Classifier for Long Method Identification

The structure of the deep neural network based classifier for long method detection is presented in Fig.1.

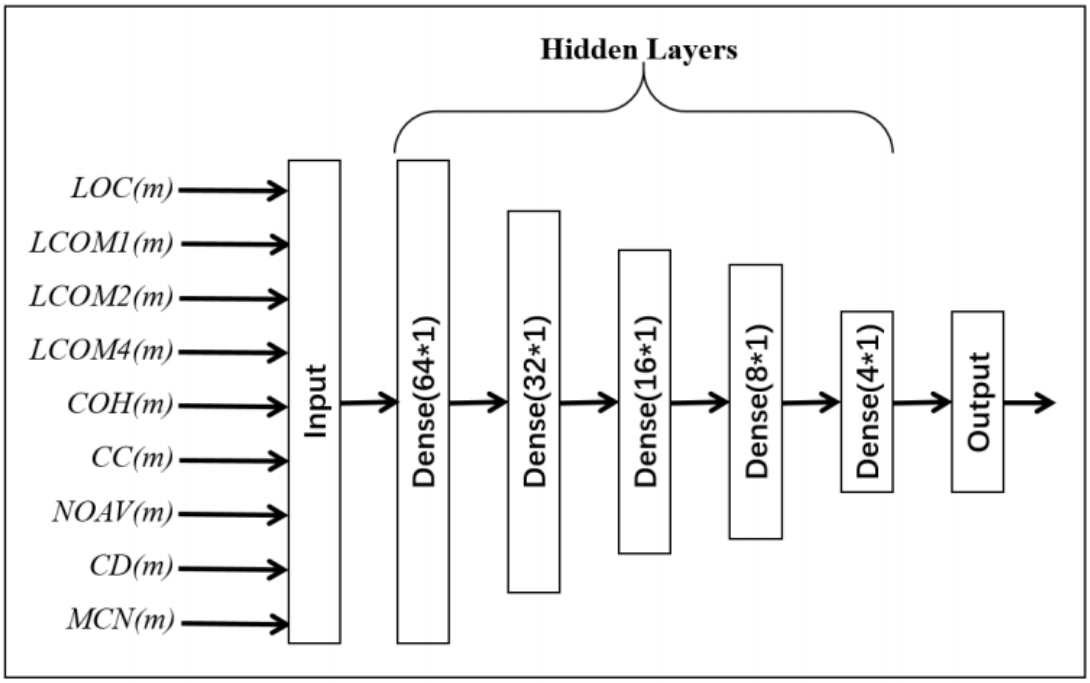

Fig.1 Classifier

The original hyperparameters of model is as follow:

- Hidden Layers Shape=(64,32,16,8,4)
- Activation Function=Relu
- Optimization Algorithm=Adam
- L2 Regularization Coefficient=0.0001

And the average performance of the model on ten projects is as follow:

Precision	Recall	F1	MCC	AUC
42.81%	78.99%	55.53%	39.53%	79.24%

In order to research the influence of different hyperparameters on model performance, we tried several combinations of parameters and compared their performance with our original parameters'. In the following tables, the first data line (bold fonts) is the performance of original parameters.

Change A Single Parameter

- **Hidden Layers Shape Modification**

Shape	Precision	Recall	F1	MCC	AUC
(64,32,16,8,4)	**42.81%**	**78.99%**	**55.53%**	**39.53%**	**79.24%**
(16,8,4)	43.77%	75.89%	55.52%	39.33%	79.21%
(64,64,64,64,64)	41.96%	78.64%	54.72%	38.30%	79.36%
(4,4,4,4,4)	42.31%	77.50%	54.74%	38.24%	79.02%
(64,32,16,8,4,4,4)	42.35%	78.14%	54.93%	38.57%	79.00%
(37,22,15,7,9)	41.98%	77.98%	54.58%	38.04%	79.35%

- **Activation Function Modification**

Activation Function	Precision	Recall	F1	MCC	AUC
Relu	**42.81%**	**78.99%**	**55.53%**	**39.53%**	**79.24%**
Identity	45.45%	62.29%	52.55%	35.40%	78.70%
Logistic	41.80%	80.43%	55.01%	38.88%	79.58%
Tanh	41.67%	78.98%	54.56%	38.07%	79.24%

- **Optimization Algorithm Modification**

Algorithm	Precision	Recall	F1	MCC	AUC
Adam	**42.81%**	**78.99%**	**55.53%**	**39.53%**	**79.24%**
LBFGS	45.61%	69.07%	54.94%	38.47%	78.78%
SGD	35.92%	80.49%	49.68%	30.47%	76.89%

- **L2 Regularization Coefficient Modification**

Coefficient	Precision	Recall	F1	MCC	AUC
0.0001	**42.81%**	**78.99%**	**55.53%**	**39.53%**	**79.24%**
0.005	41.37%	78.64%	54.22%	37.53%	78.77%
0.05	42.80%	76.76%	54.95%	38.53%	77.79%
0.5	41.73%	78.07%	54.39%	37.75%	78.85%

Changing Multiple Parameters Together

- **Coefficient=0.001 & Optimization Algorithm=Adam**
  VS
  Coefficient=0.05 & Optimization Algorithm=SGD

Precision	Recall	F1	MCC	AUC
42.81%	**78.99%**	**55.53%**	**39.53%**	**79.24%**
34.18%	56.70%	42.65%	19.72%	72.64%

- **Activation Function=Identity & Optimization Algorithm=SGD**

  The model is difficult to converge.

- **Hidden Layers Shape=(64,32,16,8,4) & Activation Function=Adam**
  VS
  Hidden Layers Shape=(4,4,4) & Activation Function=Identity

Precision	Recall	F1	MCC	AUC
42.81%	**78.99%**	**55.53%**	**39.53%**	**79.24%**
48.99%	55.26%	51.94%	35.86%	76.70%

- **Hidden Layers Shape=(64,32,16,8,4) & Activation Function=Relu & Optimization Algorithm=Adam& Coefficient=0.0001**
  VS
  Hidden Layers Shape=(64,64,64,32,16,8,4) & Activation Function=Logistic & Optimization Algorithm=SGD & Coefficient=0.05

Precision	Recall	F1	MCC	AUC
42.81%	**78.99%**	**55.53%**	**39.53%**	**79.24%**
20.58%	29.29%	24.18%	-0.05%	46.34%

- **Hidden Layers Shape=(64,32,16,8,4) & Activation Function=Relu & Optimization Algorithm=Adam**
  VS
  Hidden Layers Shape=(64,64,64,32,16,8,4) & Activation Function=Logistic & Optimization Algorithm=LBFGS

Precision	Recall	F1	MCC	AUC
42.81%	**78.99%**	**55.53%**	**39.53%**	**79.24%**
25.45%	61.24%	35.96%	4.56%	52.16%

- **Hidden Layers Shape=(64,32,16,8,4) & Activation Function=Relu & Optimization Algorithm=Adam**
  VS
  Hidden Layers Shape=(74,55,37,22,15,7,9) & Activation Function=Logistic & Optimization Algorithm=LBFGS

Precision	Recall	F1	MCC	AUC
42.81%	**78.99%**	**55.53%**	**39.53%**	**79.24%**
25.75%	83.76%	39.39%	8.56%	52.59%

- **Hidden Layers Shape=(64,32,16,8,4) & Activation Function=Relu & Optimization Algorithm=Adam**
  VS
  Hidden Layers Shape=(4,4,4) & Activation Function=Logistic & Optimization Algorithm=SGD

Precision	Recall	F1	MCC	AUC
42.81%	**78.99%**	**55.53%**	**39.53%**	**79.24%**
23.45%	51.88%	32.30%	-0.80%	54.81%