SKT-NAS: Soft Kendall's Tau based Neural Architecture Search

Di He*
South China University of Technology
Wushan 381, Tianhe District, Guangzhou, Guangdong, CN 510641
hedi2856@qq.com

Abstract

The architecture of the neural network plays a crucial role in the prediction effect of the final model. However, the traditional neural network architecture design relies heavily on expert experience and related background knowledge. And a lot of experiments and computing power consumption are required in the design process of the neural network architecture. The automatic neural network architecture search technology (NAS) based on few samples can not only get rid of the dependence on expert experience and background knowledge, but also obtain a good neural network architecture with very little computing power. RMSE or MAE is often used as the loss function for accuracy predictor training for NAS. However, the relative ranking of the performance of accurately predicting different neural network architectures in a NAS task is more important than reducing the prediction accuracy error. In this paper, a new NAS model (SKT-NAS) that focuses on the relative ranking of the performance of different neural network architectures is proposed. Non-differentiable Kendall rank correlations are designed as a class of differentiable soft forms. The NAS model is directly trained end-to-end using 1 - soft Kendall's Tau as loss. Finally, SKT-NAS achieves an average Kendall's tau of 0.79732 on 8 tasks (TOP 3) in the 2022 CVPR NAS competition track2: Performance Estimation.

1. Introduction

In recent years, neural networks have made a lot of progress and have achieved a lot of successful applications in both industry and academia [1, 2]. However, the design of the neural network structure, which plays a decisive role in the model effect, often needs to rely on the experience of human experts and a large amount of background knowledge.

Neural Architecture Search (NAS) is a method to automatically search for optimal neural network architectures without human intervention. This method gets rid of the dependence on human expert experience and background knowledge in the process of neural network structure design. NAS technology has facilitated the progress of neural networks in various fields, such as image classification [3, 4], semantic segmentation [5, 6] and super resolution [7], etc. However, the traditional NAS method needs to consume a lot of computing resources. For example, the search method based on reinforcement learning [8] needs to use 800 GPUs to search in parallel for about a month, which greatly hinders the application of NAS in practical business.

Therefore, the lightweight neural network automatic search algorithm is very meaningful. One of the methods is to train a neural network structure accuracy predictor based on a small number of samples, and then predict the performance of a large number of unknown structures [9, 10]. However, such methods usually use RMSE or MAE as the loss to train the predictor, but in the neural network architecture search task, it is more important to accurately predict the relative performance of different network structures than to reduce the difference in scores [11]. For example, the actual prediction accuracies of the two neural network architectures are 0.6 and 0.7, respectively. If RMSE or MAE is used as the loss function, the predicted loss of 0.7 and 0.8 is exactly the same as the predicted loss of 0.7 and 0.6. But obviously the next prediction reversed the order of performance of the two neural network architectures, which was a failed prediction. Therefore, training based on relative performance ranking is more reasonable.

A new NAS model (SKT-NAS) that focuses on the relative ranking of the performance of different neural network architectures is proposed in this paper. Non-differentiable rank correlation Kendall's Tau is designed to be in a differentiable form, which is called soft Kendall's Tau. 1-soft Kendall's Tau is used as the loss function to train the NAS accuracy predictor. Architectures based on Bi-LSTM with dense layers and a transformer encoder-decoder were used as the model. Finally, the model in this paper has achieved TOP3 in the 2022 CVPR NAS competition track2: Performance Estimation.

2. Method

2.1. Loss function based on rank correlation

Benefiting from back-propagation and automatic derivation, the neural network model can design various metrics as loss for end-to-end training. This paper proposes a class of differentiable loss based on rank correlation to help the NAS accuracy predictor train a better model.

Kendall's Tau (Kendall's Rank Correlation Coefficient) is a measure of the relative rank correlation between two vectors, which is often used as an evaluation metric for predictors of NAS [11, 12]. Kendall's Tau values range from -1 to 1. The larger the value of Kendall's Tau, the higher the ranking correlation between the two vectors. When the value is 1, it means that the ranking of the two vectors is exactly the same. When the value is -1, it means that the ranking of the two vectors is completely opposite. The formula for calculating Kendall's Tau is as follows:

$$Kendall's\ Tau \ = \frac{C - D}{\frac{1}{2} \times N(N-1)}$$

where:
C = the number of concordant pairs;
D = the number of discordant pairs;
N is the number of samples;

Table 1 Kendall’s Tau example

True	6	2	3	1	5	9	8	7	4	10
Pred	5	1	3	2	4	7	9	8	6	10

Table 2 Example after sorting by True

True	1	2	3	4	5	6	7	8	9	10
Pred	2	1	3	6	4	5	8	9	7	10

Taking Table 1 as an example, the calculation process of Kendall's Tau is as follows:

1) Set both variables C and D to 0;
2) Sort Pred in ascending order of True, as shown in Table 2;
3) Set the first number 2 in Pred as the comparison number, and the comparison number is compared with each number to the right of it in turn (9 times in total). If it is greater than the first number 2, the variable C+=1, and if it is less than 2, the variable D+=1. In this round, the value of C is increased by 8, the value of D is increased by 1, and C - D = 8 - 1 = 7.
4) Then set the second number 1 in Pred as the comparison number, and compare it with the 8 numbers on the right in turn. In this round, the value of C has increased by 8, and the value of D has not changed.
5) And so on, after 9 rounds, C - D = (8 - 1) + (8 - 0) + (7 - 0) + (4 - 2) + (5 - 0) + (4 - 0) - (2 - 1) + (1 - 1) + (1 - 0) = 35;
6) Finally calculate Kendall's Tau between the two vectors of True and Pred = 35 / (10 * (10 - 1) / 2) = 35 / 45 = 0.778;

In fact, n * (n - 1) / 2 of the denominator of the formula is equivalent to 9+8+7+6+5+4+3+2+1, which is equivalent to the number on the right is always greater than the comparison number in each rounds.

Obviously Kendall's Tau is a non-differentiable method. However, the comparison in the process of calculating C - D can be regarded as a sign function.

$$comparison\ result = \begin{cases} +1 & if \quad right\ number > comparison\ number \\ -1 & if \quad right\ number < comparison\ number \end{cases}$$

In each comparison, if the number on the right is larger than the comparison number, it returns +1, and if the number on the right is smaller than the comparison number, it returns -1. Then C - D can be calculated with the following formula:

$$C - D \ = \sum_{i=1}^{N} sign(Pred[i:] - Pred[i-1])$$

Although the sign function is still non-differentiable, we can use tanh (hyperbolic tangent), erf (error function) and other similar differentiable functions as an alternative to the sign function, thus Kendall's Tau is changed to a differentiable soft form (soft Kendall’s Tau).

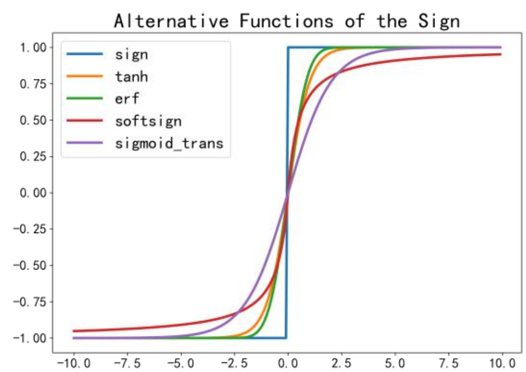

Figure 1. Alternative functions of the sign.

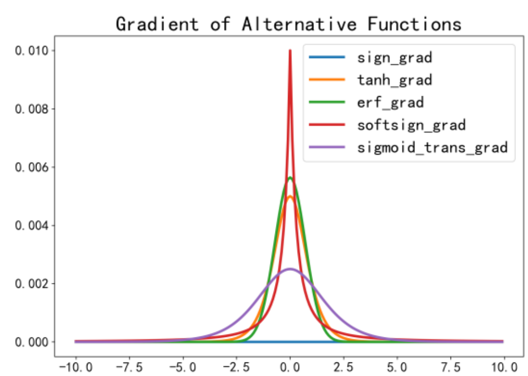

Figure 2. Gradient of alternative functions.

Figure 1 shows different alternative functions for the sign function, and Figure 2 shows the gradients of these alternative functions.

where :

softsign(x) = x / (0.5 + abs(x)),
*sigmoid_trans(x) = 2 * (sigmoid(x) - 0.5)*

Given the training principle of loss function minimization, 1 - soft Kendall's Tau is used as the loss function for model training of the NAS accuracy predictor. SKT-NAS is a class of NAS predictor models that use soft Kendall's Tau as loss.

2.2. Architecture of NAS precision predictor

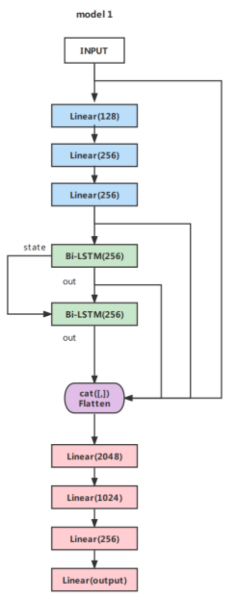

Figure 3. Architecture of model 1.

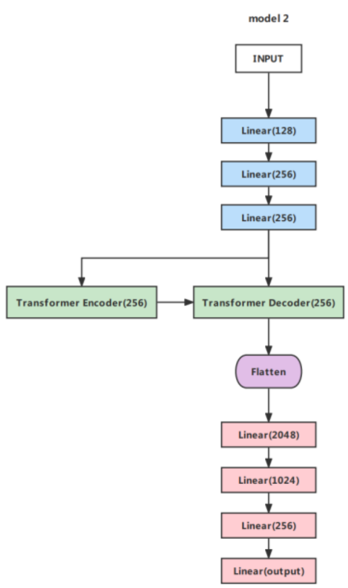

Figure 4. Architecture of model 2.

The idea of the NAS accuracy predictor in this paper is to encode the neural network architectures to be searched into a high-dimensional tensor, and then directly decode it into a relative ranking between network architectures.

Structures based on RNN transformer models are often used as encoders and decoders. There are numerous successful applications of both model architectures. The model architectures used by the SKT-NAS in this project is based on the Bi-LSTM in the RNN family, and the transformer encoder decoder. Figures 3 to 5 show the model architecture specifically used in this project. All three models use sequence data as input. In Bi-LSTM based models (Model 1 and Model 3), the original input and the output of the intermediate layer are concatenated together to form a dense layer, and finally the flattened dense layer is used with MLP to obtain the output.

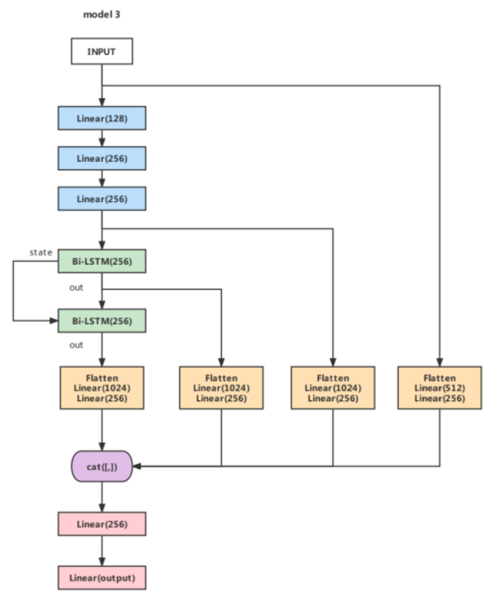

Figure 5. Architecture of model 3.

3. Experiments and results

The SKT-NAS proposed in this paper is applied to the 2022 CVPR NAS competition track2: Performance Estimation. The competition presented 100,000 completely different neural network architectures. Each neural network architecture is represented by a 37-bit character. The relative performance ranking of 500 architectures on 8 different tasks is given. The rank value is an integer from 0 to 499. Another 99,500 architectures are used as a test set, and the relative performance ranking of these architectures on 8 tasks needs to be predicted, with the value ranging from 0 to 99499. The average Kendall's Tau over the 8 tasks in the test set was used as the evaluation metric.

The experimental results are shown in Table 3.

LGB ori: original LightGBM, num_leaves=31, learning_rate=0.1;

Table 3. Comparison of different models on the test set

Model	score	cplfw	market1501	dukemtmc	msmt17	veri	vehicleid	veriwild	sop
LGB ori	0.681	0.27548	0.77417	0.77229	0.75987	0.74693	0.58998	0.78597	0.74328
Model1 RMSE	0.76799	0.23852	0.86274	0.87962	0.92935	0.88272	0.64360	0.90944	0.79795
LGB opt	0.78385	0.29680	0.87345	0.89207	0.94536	0.88662	0.66465	0.91543	0.79646
LGB opt Gaussian	0.79170	0.30304	0.88216	0.89852	0.96327	0.89229	0.66823	0.91991	0.80477
OHE Linear Gaussian	0.79251	0.30632	0.88203	0.89841	0.96363	0.89552	0.66735	0.92054	0.80633
Model3 SKT-NAS	0.79253	0.29919	0.88230	0.89703	0.96335	0.89861	0.67530	0.92263	0.80182
Model2 SKT-NAS	0.79282	0.32000	0.87896	0.89657	0.96108	0.89295	0.66939	0.92099	0.80260
Model1 pair loss	0.79395	0.29616	0.88538	0.90162	0.96621	0.90110	0.67295	0.92239	0.80579
Model1 SKT-NAS	0.79455	0.31011	0.88410	0.89642	0.96365	0.90124	0.67184	0.92352	0.80555

Model1 RMSE: Model1 with RMSELoss;

LGB opt: lightgbm after hyperparameter optimization, objective=MSE, num_leaves=2, learning_rate=0.6;

LGB opt Gaussian:LGB opt trained after labels are converted to Gaussian distribution;

OHE Linear Gaussian: Linear with one-hot input trained after labels are converted to Gaussian distribution;

Model3 SKT-NAS: Model3 with tanh soft Kendall’s Tau Loss;

Model2 SKT-NAS: Model2 with tanh soft Kendall’s Tau Loss;

Model1 pair loss: Model1 with pair loss (another relative ranking loss [11]);

Model1 SKT-NAS: Model1 with tanh soft Kendall’s Tau Loss;

The traditional machine learning model LightGBM without hyperparameter tuning performed poorly on this task with a score of only 0.681. After hyperparameter tuning, LightGBM can reach a score of 0.78385. By establishing an assumption of Gaussian distribution for the performance distribution of different neural network architectures, and using the inverse error function to convert the target to a standard normal distribution, the score can be further improved to 0.7917.

Model 1 in this paper has a score of only 0.76799 if RMSE is used as the loss function. But if the relative ranking loss is used, the scores of Model 1 can be significantly improved to 0.79395 and 0.79455, where 0.79455 uses the soft Kendall's Tau Loss, and 0.79395 uses another relative ranking loss [11]. This result proves the superiority of soft Kendall's Tau loss and relative ranking loss.

After weighted fusion of the results of lightGBM, Linear, Model1, Model2, Model3, a Leaderboard B score of 0.79732 will be obtained (TOP 3).

4. Conclusion

In this paper, the non-differentiable rank correlation Kendall's Tau is designed into a differentiable form. And a new NAS model SKT-NAS based on soft Kendall's Tau loss is proposed. Compared with MSELoss, soft Kendall's Tau Loss can bring significant improvements to the NAS model. The superior performance of SKT-NAS compared to traditional machine learning is also demonstrated.

References

[1] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[2] Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017

[3] Liu, Hanxiao, et al. "Hierarchical representations for efficient architecture search." arXiv preprint arXiv:1711.00436 (2017).

[4] Real, Esteban, et al. "Regularized evolution for image classifier architecture search." Proceedings of the aaai conference on artificial intelligence. Vol. 33. No. 01. 2019.

[5] Liu, Chenxi, et al. "Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

[6] Nekrasov, Vladimir, et al. "Fast neural architecture search of compact semantic segmentation models via auxiliary cells." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

[7] Pan, Zhihong, et al. "Real image super resolution via heterogeneous model ensemble using gp-nas." European Conference on Computer Vision. Springer, Cham, 2020.

[8] Zoph, Barret, and Quoc V. Le. "Neural architecture search with reinforcement learning." arXiv preprint arXiv:1611.01578 (2016).

[9] Deng, Boyang, Junjie Yan, and Dahua Lin. "Peephole: Predicting network performance before training." arXiv preprint arXiv:1712.03351 (2017).

[10] Sun, Yanan, et al. "Surrogate-assisted evolutionary deep learning using an end-to-end random forest-based performance predictor." IEEE Transactions on Evolutionary Computation 24.2 (2019): 350-364.

[11] Xu, Yixing, et al. "Renas: Relativistic evaluation of neural architecture search." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[12] Zheng, Xiawu, et al. "Multinomial distribution learning for effective neural architecture search." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.