Министерство образования Республики Беларусь

Учреждение образования
«Мозырский государственный педагогический университет
имени И. П. Шамякина»

ИННОВАЦИОННЫЕ ТЕХНОЛОГИИ ОБУЧЕНИЯ ФИЗИКО-МАТЕМАТИЧЕСКИМ И ПРОФЕССИОНАЛЬНО-ТЕХНИЧЕСКИМ ДИСЦИПЛИНАМ

INNOVATIVE TEACHING TECHNIQUES IN PHYSICS, MATHEMATICS, VOCATIONAL AND MECHANICAL TRAINING

Материалы XVI Международной
научно-практической конференции

Мозырь, 29 марта 2024 года

Мозырь
МГПУ им. И. П. Шамякина
2024

УДК 37:001.895
ББК 74
И66

Печатается по решению научно-технического совета
учреждения образования «Мозырский государственный
педагогический университет имени И. П. Шамякина»
(протокол № 4 от 17.04.2024)

Инновационные технологии обучения физико-математическим и профессионально-
И66 техническим дисциплинам = Innovative teaching techniques in physics, mathematics, vocational and mechanical training : материалы XVI Междунар. науч.-практ. конф., Мозырь, 29 марта 2024 г. / УО МГПУ им. И. П. Шамякина ; редкол.: И. Н. Ковальчук (отв. ред.) [и др.]. – Мозырь : МГПУ им. И. П. Шамякина, 2024. – 280 с.
ISBN 978-985-477-907-2.

В сборнике материалов конференции представлены материалы научных исследований по использованию инновационных технологий обучения физико-математическим и профессионально-техническим дисциплинам в учреждениях общего среднего, профессионально-технического, среднего специального и высшего образования.

Адресуется научным работникам, преподавателям, аспирантам, студентам.

Статьи печатаются в авторской редакции.

УДК 37:001.895
ББК 74

ISBN 978-985-477-907-2

Секция 3

Актуальные проблемы научных исследований в области физики, математики и информатики

UDC 004.852

A. D. MADAMINJONOV[1], A. E. RASHIDOV[2]
[1]Namangan state university, Namangan, Uzbekistan
[2]Samarkand state university, Samarkand, Uzbekistan

USE OF OPTIMIZERS IN MODELS

Introduction. It is well known that Optimizers help to improve the performance of models during training. Optimizers are important in deep learning to efficiently handle complex parameter spaces, speed up convergence, solve problems such as vanishing gradients, and ensure that deep neural networks generalize well to unseen data. They have a significant impact on the effectiveness and efficiency of training deep learning models [1-6].

In this study, we investigated the effect of different optimizers on the accuracy of our preprocessed image data model.

Mehmood F., Ahmad, S., and Whangbo, T. K.'s article "Efficient Optimization Techniques for Training Deep Neural Networks" also provides a lot of useful information about optimizers. In this study, the main task was to improve the optimization technique for training deep neural networks. Deep learning has played a major role in computer vision tasks such as image classification, natural language processing, and object recognition. In deep neural network training, optimizers have been shown to be important in increasing accuracy and reducing training time. Optimizers are algorithms used to improve the efficiency of the model or minimize the loss function. Recent research suggests many state-of-the-art optimizers to support neural network training. Each optimizer has its advantages and disadvantages. This study focuses on various optimization methods and aims to introduce an optimization method to improve the accuracy of deep neural networks. The study experimented with CIFAR-10 and CIFAR-100 datasets using VGG16, ResNet and DenseNet models. The study made some changes to the Adam optimizer in the algorithm by minimizing the cyclic path, removing the extra hyperparameter, and changing the position of epsilon. The researchers used SGD, Adam and RMSProp optimizers and compared the training and testing accuracy with the proposed optimization method. They concluded that the proposed methodology outperforms current state-of-the-art optimizers with marginally improved accuracy. The proposed optimizer achieved 97.98% accuracy in training and 95.95% accuracy in testing. They also observed that the training accuracy of the proposed optimization method was slightly lower than that of the RMSProp optimizer when using the VGG16 model. To conclude, the overall results show that the proposed method is effective in achieving accuracy and works well with the current architecture [5, 11].

Deep learning is a sub-field of machine learning that is used to perform complex tasks such as speech recognition, text classification, etc. A deep learning model consists of activation function, input, hidden, output layers, loss function, etc. All known deep learning algorithms attempt to summarize data using an algorithm and make predictions based on unseen data. We need an algorithm that compares input instances to outputs along with an optimization algorithm. The optimization algorithm finds the value of the parameters (weights) that minimizes the error when comparing the inputs to the outputs. This article provides an overview of such optimization algorithms or optimizers in deep learning [1].

In deep learning, optimizers are algorithms that adjust model parameters during training to minimize the loss function. They allow neural networks to learn from data by iteratively updating weights and uncertainties. Common optimizers include Stochastic Gradient Descent (SGD), Adam, and RMSprop. Each optimizer has specific update rules, learning rate, and momentum to find optimal model parameters to improve performance [7].

Epoch - How many times the algorithm runs on the entire training data set

A sample is a single row of a data set. Batch refers to the number of samples to be taken to update the model parameters. Learning rate is a parameter that tells the model how many model weights to update. Loss function - a cost function is used to calculate the cost, which is the difference between the estimated value and the actual value. Weights/bias - Learned parameters in the model that control the signal between two neurons.

Gradient Descent. Gradient descent is an optimization algorithm based on a convex function that iteratively changes its parameters to reduce a given function to a local minimum. Gradient descent moves in the opposite direction to the steepest ascent, iteratively decreasing the loss function. Finding minima is a process that depends on the derivatives of the loss function. Memory-intensive and process-slowing parameters require the use of the entire training set to compute the gradient of the cost function.

Stochastic gradient descent SGD is an iterative method for optimizing an objective function with suitable smoothness properties (differentiable or subdifferentiable). This can be viewed as a stochastic approximation of gradient descent optimization, as it replaces the true gradient with its estimate. It reduces computational complexity, especially in high-dimensional optimization problems, and achieves faster iterations at the cost of lower convergence rates [8, 9].

Mini-batch gradient descent is a variation of the gradient descent algorithm that divides the training data set into small chunks that are used to calculate model error and update model coefficients [10]. Small-batch gradient descent seeks to find a balance between the robustness of stochastic gradient descent and the efficiency of batch gradient descent. It is one of the most common implementations of gradient descent used in the field of deep learning [6].

Momentum SGD is a stochastic optimization method that adds a momentum term to regular stochastic gradient descent. Momentum simulates the inertia of an object as it moves, meaning that the previous update direction is somewhat preserved during the update, while the current update gradient is used to adjust the final update direction. So we can increase stability to some extent to learn faster and also get rid of local optimization [12, 13]. The learning rate is constant in all algorithms, and the intuition behind AdaGrad is that we can use different learning rates for each hidden layer and for each neuron based on different iterations. RMS-Prop is a special version of Adagrad where the learning rate is an exponential average of the gradients instead of the sum of the squared gradients. RMS-Prop essentially combines momentum with AdaGrad.

Adadelta is an enhanced version of Adagrad that attempts to reduce Adagrad's monotonically learning rate and overcome the decreasing learning rate problem. Adadelta takes the ratio of the running average of previous time steps to the current gradient.

The Adam optimizer is one of the most popular gradient descent optimization algorithms. This is a method that calculates the adaptive learning rate for each parameter. It preserves the decaying average of past gradients similar to momentum, the decaying average of past squared gradients similar to RMS-Prop and Adadelta, in short it combines the advantages of both methods.

Conclusion

In conclusion, among the optimizers that are of major importance in deep learning, currently Adam optimizer shows the best accuracy in satisfactory time, RMSprop can show accuracy similar to Adam's accuracy, but requires much more computation time. The SGD algorithm takes the least amount of time to train and is one of the best performing optimizers. But to achieve the accuracy of Adam's optimizer, SGD requires more iterations and hence the computation time increases. Pulsed SGD has an unexpectedly large computational time and shows almost the same accuracy as SGD. This means that the obtained momentum value should be optimized. We can get these insights through the results obtained in the table below.

Epoch	SGD	Ada_Grad	RMS Prop	AdaDelta	Adam
Epoch100	0.52	0.96	0.94	0.97	0.96
Epoch200	0.87	0.98	0.94	0.97	0.97
Epoch300	0.92	0.98	0.95	0.98	0.97
Epoch400	0.94	0.98	0.95	0.98	0.98
Epoch500	0.95	0.98	0.96	0.98	0.98
Epoch600	0.96	0.98	0.94	0.98	0.98
Epoch700	0.96	0.98	0.97	0.98	0.98
Epoch800	0.96	0.98	0.95	0.98	0.98
Epoch900	0.97	0.98	0.96	0.98	0.98
Epoch1000	0.97	0.98	0.97	0.98	0.98

References

1. Gupta, A. (2023, September 13). A comprehensive guide on optimizers in deep learning. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-deep-learning-optimizers/.

2. Axatov A.R., Rashidov A.E. "Improving the efficiency of intellectual analysis using the correlation coefficient" "Raqamli ta'lim texnologiyalari: amaliyot, tajriba, muammo va istiqbollari" mavzusidagi Respublika ilmiy-amaliy anjumani, 2023, 4–5 – iyul, 213–216-bet.

3. Akhatov A., Rashidov A., Renavikar A. "Optimization of the database structure based on Machine Learning algorithms in case of increased data flow" International Conference on Artificial Intelligence, Blockchain, Computing and Security (ICABCS 2023), February 24–25, 2023, 675–681.

4. Musstafa. (2022, February 12). Optimizers in Deep Learning – MLearning.ai – medium. Medium. https://medium.com/mlearning-ai/optimizers-in-deep-learning-7bf81fed78a0.

5. Axatov A.R., Rashidov A.E., Nazarov F.M., Renavikar A. "Optimization of the number of databases in the Big Data processing" Проблемы информатики, № 1 (58), 2023, DOI: 10.24412/2073-0667-2023-1-33-47.

6. Brownlee, J. (2019, August 19). A gentle introduction to Mini-Batch Gradient descent and how to configure batch size. MachineLearningMastery.com. https://machinelearningmastery.com/gentle-introduction-mini-batch-gradient-descent-configure-batch-size/.

7. Bushaev, V. (2018, June 21). Stochastic Gradient Descent with momentum - Towards Data Science. Medium. https://towardsdatascience.com/stochastic-gradient-descent-with-momentum-a84097641a5d.

8. Axatov A.R., Rashidov A.E., Nazarov F.M. "Real-Time Big Data Processing Based on a Distributed Computing Mechanism in a Single Server" Stochastic Processes and Their Applications in Artificial Intelligence. IGI Global, May, 2023, DOI: 10.4018/978-1-6684-7679-6.ch009.

9. Stochastic gradient descent / https://en.wikipedia.org/wiki/Stochastic_gradient_descent.

10. Rashidov A., A.Akhatov, F.Nazarov "The Same Size Distribution of Data Based on Unsupervised Clustering Algorithms" ICAILE 2023: Advances in Artificial Systems for Logistics Engineering III. (LNDECT, volume 180), 437–447 p, doi: https://doi.org/10.1007/978-3-031-36115-9_40.

11. Mehmood, F., Ahmad, S., & Whangbo, T. K. (2023). An efficient optimization technique for training deep neural networks. Mathematics, 11(6), 1360. https://doi.org/10.3390/math11061360

12. A.R.Akhatov, Rashidov A.E. "Machine Learning asosida taqsimlangan ma'lumotlar bazasi sonini optimallashtirish" "Kompyuter ilmlari va muhandislik texnologiyalari" Xalqaro ilmiy texnik konferensiya, 2022, 14–15 oktabr, 110–112

13. A. Rashidov, A. Akhatov, I. Aminov, D. Mardonov "Distribution of data flows in distributed systems using hierarchical clustering" International conference on Artificial Intelligence and Information Technologies (ICAIIT 2023), Samarkand 2023, Novamber 3–4.

UDC 004.852

A. D. MADAMINJONOV
Namangan state university, Namangan, Uzbekistan

ANALYSIS OF METHODS AND ALGORITHMS OF DETERMINING PERSONAL CHARACTERISTICS

Introduction. With the continuous development of computer technology, human dependence on network technology has increased, which has increased the importance of security issues. User authentication is an important tool to protect against security vulnerabilities and attacks. Different authentication methods (fingerprint scanning, voice recognition, methods such as SMS, one-time passwords and facial recognition) are used. Face recognition is one of the important applications for processing still images and videos. Creating an automated system equivalent to human face recognition is a complex process [1].

There are many ways to detect a face. Viola-Jones algorithm, knowledge or rule-based, feature-based, template matching, appearance-based, convolutional neural network-based unified gravity detector, etc. [3]. Under ideal conditions, facial recognition systems can have near-perfect accuracy. Using facial recognition algorithms (such as a searchable person's photo or passport photo) can achieve 99.97% accuracy in standard situations. This can be compared to Iris scanners. This type of facial verification has become so reliable that even banks are using this system to log users into their accounts.

However, this level of accuracy is only possible under ideal conditions where there is consistency in lighting and positioning and the subjects' facial features are clear and unobstructed. In the real world, the level of accuracy is much lower. For example, FRVT found that one leading algorithm had an error rate of less than 0.1% compared to a human image in high-quality search, images of individuals taken "in the wild" where the subject may not be looking directly o increased to 9.3% when matched instead. The camera can be obscured by objects or shadows, which is

Секция 3

Актуальные проблемы научных исследований в области физики, математики и информатики