ISSN: 2664-9640

The future of machine learning and algorithms

Kuan Xiong [a,*] Mingxuan Li[b]

Central South University, Hunan ChangSha ,410000, China

[a]15040340137@xs.hnit.edu.cn, [b]limingxuancn@163.com

Abstract

Machine learning is a science of artificial intelligence, and the main research object in this field is artificial intelligence. Machine learning refers to the use of computers to learn the inherent regularity information in data, to obtain new experience and knowledge, to improve the intelligence of computers, and to enable computers to make decisions like humans. With the increasing demand for data in various industries, the efficiency of processing and analyzing data has become higher, and a series of machine learning algorithms have emerged. Machine learning algorithms mainly refer to the steps and processes of solving optimization problems through mathematical and statistical methods. For different data and different model requirements, selecting and using appropriate machine learning algorithms can solve some practical problems more efficiently.

Keywords

Machine learning; deep learning; artificial intelligence; optimization algorithm.

1. The history of machine learning

The development of machine learning and the development of artificial intelligence are inseparable, and machine learning is an inevitable product of the development of artificial intelligence research to a certain stage. Machine learning can be traced back to the study of artificial neural networks. In 1943, literature [1] and literature [2] proposed the neural network hierarchical structure model, established the computational model theory of neural network, and laid the foundation for the development of machine learning. In 1950, Turing, the "father of artificial intelligence", proposed the famous "Turing Test", which made artificial intelligence an important research topic in the field of science.

Literature [3] proposed the concept of Perceptron, and for the first time used an algorithm to accurately define the neural network mathematical model of self-organization and self-learning, and designed the first computer neural network. This machine learning algorithm became the originator of the neural network model. In 1959, IBM designed a checkers program with learning ability, which once defeated the undefeated champion of the United States for 8 years [4]. This program gave people an initial demonstration of the power of machine learning. Literature [5] found that the unique neural network structure in cat cerebral cortex can effectively reduce the complexity of learning, and thus proposed the famous Hubel-Wiese biological vision model. The pioneer of artificial intelligence research published the book "Perceptron" [6], which has far-reaching influence on machine learning research, in which the assertion of the basic idea of machine learning: the algorithm ability and computational complexity of solving problems has far-reaching influence and continues to this day.

In the summer of 1980, the first International Symposium on Machine Learning was held at Carnegie Mellon University in the United States, marking the rise of machine learning research worldwide. In 1986, "Machine Learning" was founded, marking that machine learning gradually attracted the attention of the world and began to accelerate its development.

In 1986, the literature [7] jointly published the famous back-propagation algorithm (BP) in the journal "Nature". In 1989, American Bell Labs scholars [8] proposed the most popular convolutional neural network (CNN) computing model, derived an efficient training method based on the BP algorithm, and successfully applied it to English handwriting recognition.

In the 1990s, multiple shallow machine learning models came out one after another, such as logistic regression, support vector machines, etc. The commonality of these machine learning algorithms is that the mathematical model is the optimization problem of a convex cost function, the theoretical analysis is relatively simple, and it is easy to obtain from the training samples. Learn the internal mode to complete the primary intelligent work such as object recognition and character assignment.

In 2006, the leading literature in the field of machine learning [9] published an article and proposed a deep learning model. The main arguments include: artificial neural network with multiple hidden layers has good feature learning ability; through layer-by-layer initialization to overcome the difficulty of training and achieve overall network tuning. The proposal of this model opens a new era of deep network machine learning. In 2012, the literature [10] research team used the deep learning model to win the most influential ImageNet competition in the field of computer vision, marking the second stage of deep learning.

Deep learning has achieved impressive results in many fields in recent years, and has launched a number of successful commercial applications, such as Google Translate, Apple's voice tool Siri, Microsoft's Cortana personal voice assistant, and Ant Financial's Smile to Pay. technology. Especially in March 2016, Google's AlphaGo and Go world champion and professional nine-dan player Lee Sedol competed in the human-machine battle of Go, winning by a total score of 4 to 1. On October 18, 2017, the DeepMind team announced the strongest version of AlphaGo, code-named AlphaGo Zero, which can learn from a blank state without any human input, and the self-training time is only 3 days. The number is 4.9 million sets, and it can beat the seniors with a 100:0 record.

In general, the "machine learning period" is divided into three stages. In the 1980s, connectionism was more popular, representing work with perceptrons and neural networks. In the 1990s, statistical learning methods began to occupy the mainstream stage, and the representative method was support vector machine. In the 21st century, deep neural networks were proposed, and connectionism has always been on the rise. With the continuous improvement of data volume and computing power, deep learning is used as a Many basic AI applications gradually mature in literature [11].

2. Common algorithms for machine learning

Machine learning algorithms can be divided into 3 types: supervised learning, unsupervised learning, and reinforcement learning. Semi-supervised learning can be considered as a combination of supervised learning and unsupervised learning, so it is not within the scope of this article for the time being.

Supervised learning learns a model by training samples, and then uses this model for inference. For example, if we want to recognize images of various fruits, we need to use samples labeled manually (that is, labeled with the category to which each image belongs, such as apples, pears, bananas) for training to get a model. Next, we can Using this model to make judgments about unknown types of fruit is called prediction. If you only predict a category value, it is called a classification problem; if you want to predict a real number, it is called a regression problem, such as predicting a person's income based on a person's education, years of work, city, industry and other characteristics. Typical algorithms are: Support Vector Machines, Decision Trees, Naive Bayesian Classification and K-Nearest Neighbors [12].

ISSN: 2664-9640

Unsupervised learning does not have a training process. Given some sample data, let the machine learning algorithm directly analyze the data to obtain some knowledge of the data. The typical representative is clustering. For example, we crawl 10,000 web pages, and we want to complete the classification of these web pages. Here, we do not have pre-defined categories, nor do we have a trained classification model. The clustering algorithm needs to complete the classification of these 10,000 web pages by itself, to ensure that the same type of web pages have the same theme, and different types of web pages are different. Another type of typical algorithm for unsupervised learning is data dimensionality reduction, which transforms a high-dimensional vector into a low-dimensional space and maintains some intrinsic information and structure of the data. Typical algorithms are: principal component analysis, singular value decomposition and K-means clustering.

Reinforcement learning is a special type of machine learning algorithm. The algorithm determines an action to execute according to the current environmental state, and then enters the next state. The goal is to maximize the benefits obtained. For example, the game of Go is a typical reinforcement learning problem. At each moment, it is necessary to decide where to drop the chess according to the current chess position, and then proceed to the next state, repeatedly placing chess pieces until the game is won or lost. The goal here is to win as many games as possible to get the maximum reward. A typical algorithm is Q-learning.

In summary, classification algorithms solve the "what" problem, that is, predicting the class to which a sample belongs. Regression algorithms solve the "how much" problem, that is, predict a quantitative value from a sample. The clustering algorithm solves the problem of "how to classify", ensuring that the samples of the same class are similar, and the samples of different classes are as different as possible. Reinforcement learning solves the "how to" problem, that is, decides what action to perform based on the current state, and finally gets the maximum reward [13].

3. Commonly used evaluation indicators

In the field of artificial intelligence, the effect of machine learning needs to be evaluated by various indicators [14]. The evaluation index of the model is used to evaluate the quality of the model training. Machine learning performance evaluation criteria are the premise of model optimization. In the process of designing machine learning algorithms, different problems require different evaluation criteria. This chapter summarizes the commonly used indicators of machine learning algorithms.

Consider a binary problem, that is, classifying instances into positive or negative classes. For a dichotomous problem, four situations arise. If an instance is a positive class and is also predicted to be a positive class, it is a true class, and if an instance is a negative class and is predicted to be a positive class, it is called a false positive class. Correspondingly, if the instance is a negative class is predicted to be a negative class, it is called a true negative class, and a positive class is predicted to be a negative class, it is a false negative class.

TP: the number of correct positives;

FN: false negatives, the number of matches not found correctly;

FP: false positive, the match given is incorrect;

TN: number of non-matching pairs correctly rejected;

Introduce two new nouns from the contingency table. One is the true class rate,

The calculation formula is

$$TPR = TP / (TP + FN)$$

It depicts the proportion of positive instances identified by the classifier to all positive instances. The other is the negative and positive class rate, which is calculated as

FPR = FP / (FP + TN)

What is calculated is the proportion of all negative instances that the classifier mistakenly believes to be a positive class. There is also a true negative class rate, also known as specificity, which is calculated as

TNR = TN /(FP + TN) = 1 - FPR

Precision and recall are two metrics widely used in the fields of information retrieval and statistical classification to evaluate the quality of results. The precision is the ratio of the number of relevant documents retrieved to the total number of documents retrieved, which measures the precision rate of the retrieval system; the recall rate refers to the ratio of the number of relevant documents retrieved to the number of all relevant documents in the document library, which measures the is the recall rate of the retrieval system.

In general, Precision is how many of the retrieved items are accurate, and Recall is how many accurate items are retrieved. The definitions of the two are as follows:

Precision = number of correct messages extracted / number of messages extracted

Recall = Number of correct messages extracted / Number of messages in the sample

In order to evaluate the pros and cons of different algorithms, the concept of F1 value is proposed on the basis of Precision and Recall to evaluate Precision and Recall as a whole. F1 is defined as follows:

F1 value = correct rate * recall rate * 2 / (correct rate + recall rate)

Of course, it is hoped that the higher the precision of the retrieval result, the better, and the higher the Recall, the better, but in fact, the two are contradictory in some cases. For example, in extreme cases, we only search for one result, and it is accurate, then the Precision is 100%, but the Recall is very low; and if we return all the results, for example, the Recall is 100%, but the Precision will be very low. Therefore, in different occasions, you need to judge by yourself whether you want the Precision to be higher or the Recall to be higher. If you are doing experimental research, you can draw a Precision-Recall curve to help with analysis.

Precision and Recall indicators sometimes contradict each other, so they need to be considered comprehensively. The most common method is F-Measure (also known as F-Score).

F-Measure is Precision and Recall weighted harmonic mean:

$$F = \frac{(\alpha^2 + 1)P * R}{\alpha^2 (P + R)}$$

When the parameter α=1, it is the most common F1. Therefore, F1 combines the results of P and R, and when F1 is higher, the test method is more effective.

In a binary classification model, for the continuous results obtained, it is assumed that a threshold has been determined, such as 0.6. Instances greater than this value are classified as positive, and instances smaller than this value are classified as negative. If the threshold is reduced to 0.5, of course, more positive classes can be identified, that is, the ratio of identified positive examples to all positive examples is increased, that is, TPR, but at the same time, more negative examples will be treated as A positive example is made, that is, the FPR is improved. To visualize this change, ROC is introduced, and the ROC curve can be used to evaluate a classifier.

ROC (Receiver Operating Characteristic) is translated as "Receiver Operating Characteristic Curve". The curve is drawn by two variables, 1-specificity and Sensitivity. 1-specificity=FPR, which is the negative and positive class rate. Sensitivity is the true class rate, TPR (True positive rate), which reflects the degree of positive class coverage. This combination is 1-specificity to sensitivity, that is, cost to benefit.

The AUC value is the area covered by the ROC curve. Obviously, the larger the AUC, the better the classification effect of the classifier.

AUC = 1, which is a perfect classifier. When using this prediction model, no matter what threshold is set, a perfect prediction can be obtained. In the vast majority of prediction cases, there is no perfect classifier.

0.5 < AUC < 1, better than random guessing. This classifier (model) can have predictive value if the threshold is properly set.

AUC = 0.5, the following machine guesses the same (for example: losing a copper plate), the model has no predictive value.

AUC < 0.5, worse than random guessing; but better than random guessing as long as it always works against predictions.

The physical meaning of AUC: Assuming that the output of the classifier is the socre (confidence) that the sample belongs to the positive class, the physical meaning of AUC is that, taking any pair of (positive and negative) samples, the score of the positive sample is greater than the score of the negative sample. probability.

4. Outlook

At present, the great progress in the research and application of machine learning represented by deep learning is obvious to all, which has effectively promoted the development of artificial intelligence. However, it should also be noted that it is still a new thing after all, and most of the conclusions are obtained through experiments or experience, which still needs to be further studied and supported by theory. Yann LeCun, a professor at New York University and one of the founders of CNN, pointed out several key limitations of deep learning at the 2015 IEEE Conference on Computer Vision and Pattern Recognition: lack of theoretical basis and reasoning mechanism behind the work; lack of short-term memory; inability to perform Unsupervised learning. Deep learning based on multi-layer artificial neural network is inspired by the layered work of human cerebral cortex. Although deep learning is currently the closest intelligent learning method to the human brain, the current deep network is similar to the human brain in structure, function and mechanism. larger gap. In addition, there is still a lack of accurate cognition of the structure and mechanism of the cerebral cortex itself. If it is to truly simulate the nervous system composed of more than 10 billion neurons in the human brain, it is still difficult to achieve. Therefore, research on computational neuroscience also needs to go a long way. In addition, the network structure, algorithms and parameters of machine learning models are becoming larger and more complex. Usually, accurate models can only be trained with the support of large amounts of data and large amounts of computation. The requirements for the operating environment are getting higher and higher, and the resources are occupied more and more. The more, this also raises its application threshold. Undoubtedly, machine learning, as an important branch of artificial intelligence, has made great progress in many fields and shows strong development potential. However, it should be noted that the development of machine learning is still in the initial stage. Although there are various machine learning algorithms, they cannot fundamentally solve the barriers faced by machine learning. Machine learning still mainly relies on supervised learning, and has not yet overcome Weak artificial intelligence. So we still have a long way to go for machine learning. In short, machine learning is in the ascendant and has broad research and application prospects, but the challenges it faces cannot be ignored. Only the dialectical unity of the two can push machine learning to a higher level.

References

[1] Yang Hao,Li Jiaxi,Liu Siru,Yang Xiaoling,Liu Jialin. Predicting Risk of Hypoglycemia in Patients With Type 2 Diabetes by Electronic Health Record-Based Machine Learning: Development and Validation.[J]. JMIR medical informatics,2022,10(6).

[2] Yang Hao,Li Jiaxi,Liu Siru,Yang Xiaoling,Liu Jialin. Predicting Risk of Hypoglycemia in Patients With Type 2 Diabetes by Electronic Health Record-Based Machine Learning: Development and Validation.[J]. JMIR medical informatics,2022,10(6).

[3] Chatterjee Ankita,Saha Jayasree,Mukherjee Jayanta. Clustering with multi-layered perceptron[J]. Pattern Recognition Letters,2022,155.

[4] Tilman Mehler. Dynamic Incremental Hashing in Program Model Checking[J]. Electronic Notes in Theoretical Computer Science,2006,149(2).

[5] Li Bin,Todo Yuki,Tang Zheng. Artificial Visual System for Orientation Detection Based on Hubel–Wiesel Model[J]. Brain Sciences,2022,12(4).

[6] Yu Yuxiu. Application of the Neural Network Based on the Multilayer Perceptron Genetic Algorithm in Chinese-English Two-Way Translation[J]. Journal of Sensors,2022,2022.

[7] Singh Abha,Kushwaha Sumit,Alarfaj Maryam,Singh Manoj. Comprehensive Overview of Backpropagation Algorithm for Digital Image Denoising[J]. Electronics,2022,11(10).

[8] Ay Betul,Turker Cihan,Emre Elif,Ay Kevser,Aydin Galip. Automated classification of nasal polyps in endoscopy video-frames using handcrafted and CNN features[J]. Computers in Biology and Medicine,2022,147.

[9] Oh Jang-Hoon,Kim Hyug-Gi,Lee Kyung Mi,Ryu Chang-Woo. Reliable quality assurance of X-ray mammography scanner by evaluation the standard mammography phantom image using an interpretable deep learning model[J]. European Journal of Radiology,2022,154.

[10] Alsudais, Abdulkareem. Extending ImageNet to Arabic using Arabic WordNet[J]. Multimedia Tools and Applications,2022(prepublish).

[11] Roster Kirstin,Connaughton Colm,Rodrigues Francisco A.. Forecasting new diseases in low-data settings using transfer learning[J]. Chaos, Solitons and Fractals: the interdisciplinary journal of Nonlinear Science, and Nonequilibrium and Complex Phenomena,2022,161.

[12] Ye Yulong,Li Lingjie,Lin Qiuzhen,Wong Ka-Chun,Li Jianqiang,Ming Zhong. Knowledge guided Bayesian classification for dynamic multi-objective optimization[J]. Knowledge-Based Systems, 2022,250.

[13] Rajabi Enayat,Sahebari Maryam,Thomas Tressy. Analyzing systemic lupus erythematosus publications using neural network-based multi-label classification algorithms.[J]. Lupus, 2022,31(7).

[14] Nasir Nida,Kansal Afreen,Alshaltone Omar,Barneih Feras,Sameer Mustafa,Shanableh Abdallah,Al-Shamma'a Ahmed. Water quality classification using machine learning algorithms[J]. Journal of Water Process Engineering,2022,48.