Computer Life
ISSN: 1819-4818 | Vol. 12, No. 1, 2024

The Applications of Machine Learning Algorithms to the Field of Medical Diagnosis

Zihan Yin

School of Pharmacy, Xuzhou Medical University, Pharmaceutical Preparation Major, Nanjing ,211200, China

Abstract: With a large population base, the medical needs are also very large. Patients need a lot, and doctors have a big workload. In this case, the pressure of doctors increases accordingly. The competence and clinical experience of the doctor are at the heart of medical diagnosis. Accumulating experience requires massive practice and literature reading and comprehension. With the development of computer science and the deep study of machine learning theory, machines are gradually involved in the process of medical diagnosis. The application of machine learning algorithms has opened up a path for medical diagnosis, which can assist doctors to improve the efficiency of diagnosis and treatment. The main content of this paper is the main problems and their related machine learning algorithms, the challenges and future development.

Keywords: Machine learning, medical diagnosis， Support vector meachine, Neural network.

1. Introduction

The first part of this paper is the research background and related machine learning algorithm theory, the second part is the main research problems, and the third part is the challenges and development trends.In modern society, people's increasing pressure leads to irregular work and rest, unbalanced diet distribution, and the decline of health status and increased medical demand. And health is extremely important to a person. The background and purpose of this study is to increase medical demand and increase work volume, thus reduce efficiency and increase error rate.

The main research problems can be divided into medical imaging problems and non-image problems, mainly through chemical detection methods. This is the basic means of medical diagnosis.

2. Main problems and related machine learning algorithms

2.1. Diagnosis Challenges faced

Doctors need sufficient professional knowledge and practice in the process of diagnosis, and people will forget it, and need a long time of learning and practice to maintain a sufficient level. And one can't finish a very high amount of work endlessly. It takes a long time to train a doctor, so it is not very efficient. And the machine can work tirelessly, under the training of the machine learning algorithm can greatly improve the work efficiency, is a more appropriate idea.

2.2. Machine Learning Solutions

2.2.1. Support Vector Machine (SVM)

The SVM model is a classification algorithm that can classify diseases into different labels, allowing the machine to participate in the diagnosis after learning and optimization. The algorithm learns the training sample set D= $\{(x_1, y_1), (x_2, y_2),(x_m, y_m)\}$, y $\{-1, +1\}$, mainly to find an appropriate boundary in the sample space to separate different categories of samples. For a variety of different diseases, one can be used as positive cases and the rest as counter examples. Each category of disease was studied as a positive case. In the sample space, the linear equation $W^TX + b=0$ can be used as a representation to divide the hyperplane. Suppose the hyperplane (W, b) can correctly classify the training sample, for (x_i, y_i), $y_i(W^TX+b)>=1$ Sample points in the sample set with a discriminant equal to 1 are called the support vector. The SVM accurately identifies diseases by classifying the resulting chemical detection data. In the actual detection process, not every case can be completely linearly segmented, we can convert it into high dimensional space for linear segmentation; similarly, soft interval support vector machine model: $y_i(W^t X + b)> = 1-s_i$. $s_i>=0$,i=1,2,,m Above, s_i is the relaxation variable. The relaxation variable is introduced by allowing some samples that do not satisfy the constraints to avoid overfitting.

2.2.2. Convolutional Neural Network

Convolutional neural network realizes the learning of target image features by simulating the cognitive process of human brain. It is an evolutionary version of the multi-layer perceptron. The data operation process is more complex than the perceptron. It mainly consists of three basic structures: convolutional layer, pooling layer and full connection layer. The network divides the image into many pixels, selects it by the convolution layer through the convolution core, calculates the convolution core with the selected image data, and outputs the results. The pooling layer is like using a "pool" to select the data, which serves to downsample the output of the convolution layer to reduce the computation. The last output label.

2.2.3. Noise Reduction Algorithm

For a learner, the tested samples are often more complex than the training samples and will have more noise interference judgment. Learners can't completely avoid noise to be completely correct, and we humans can't be completely correct. There is a "no free lunch" law that the expected performance of the learners is irrelevant to the algorithm itself. The learner can only constantly approach the optimal performance, just like the intersection of the two cut lines, and the tangent line is like the optimal performance. Noise is not the target component we have learned. To reduce noise, principal component analysis can be used to reduce the dimension, and the weight of the feature dimension of interference can be reduced or even reduced to 0. Deep learning model can also be used to reduce the weight of

interference dimension and use deep neural network for accurate diagnosis.

3. Application in specific diagnostic problems

3.1. Medical Image Analysis

Diagnosis of disease by medical imaging can be regarded as identifying all physiological tissues present in the image, whether and not each tissue is diseased, what kind of lesion belongs to, and all circumstances in which the lesion may occur. The convolutional neural network is used to identify the pathological tissue in the image, extract the feature value of the identified image, calculate and discriminate under the framework of support vector machine algorithm, and obtain the analysis results.

3.2. Non-image Pathology Classification

For the diagnostic means of chemical detection, the disease can be accurately classified by measuring the content of biomarkers and analyzing the characteristic value. The eigenvalues from biotable quantification were fed into the network calculation output. For the symptoms of patients, the feature value was extracted through the speech recognition network, and the classification task was conducted to judge the type.

4. Challenges in Machine Learning for Medical Diagnosis

Despite the notable achievements of machine learning algorithms in the field of medical diagnosis, new challenges remain. For example, data quality and annotation problems: the quality of medical data and its annotation are the key factors in diagnosis and treatment, and more effective data preprocessing and annotation technology are needed to achieve better diagnosis and treatment results.Insufficient amount of data will affect the learning effect, but medical data often involves patients' privacy, and its privacy protection is also an important issue.To ensure fairness and transparency and avoid discrimination and prejudice, it also needs to be properly handled.

Different data sources, their data formats and judgment criteria may be different, and the workload of standardized processing is also very large. Therefore, the cost of obtaining the training data is relatively high independently. It is not completely reasonable to learn completely based on the annotated data. For the test samples with large differences in feature values but the same label, the learner can make a correct judgment. This requires training using unsupervised learning theory. Unsupervised learning is broad, and how to learn the required abilities is a problem that needs to be addressed.

5. Future development trend

With the continuous progress and development of science and technology, the future algorithm diagnosis will achieve higher accuracy requirements: Improve the efficiency of the diagnostic process: With the development and promotion of intelligent health system, the monitoring of human health status, recording data analysis and health reminder can enhance health awareness, prevent the deterioration of diseases, and maintain people's health.

In addition to the accuracy requirement, the popularity of related diseases will also increase. The current hot issues are lung nodules, breast cancer, CT and rectal imaging, and prostate cancer. Under the influence of living habits, some unappreciated diseases such as diabetes begin to seriously harm people's health. These diseases may become a new research hotspot in the future.

The interpretability of machine learning models is also important issues. Doctors do not understand the thinking of the machine, so they cannot judge the outcome of the machine, so the diagnosis efficiency is reduced.

6. Conclusion

Machine learning algorithms have been significantly applied in medical diagnosis. Their goal is to improve the generalization ability of machine learning algorithms, improve the level of medical services, and improve people's quality of life. In my opinion, in the future development, how to find and solve the overfitting is the most important and also the most basic problem, which is closely related to the data. In this information society, the role of data is very important. As a public manager, the government should ensure data security and protect personal privacy. The use of data should be reasonable in all industries to avoid the confusion of the classified data. Everyone should protect their privacy and not do anything like leaking confidential data.

In the future, with the continuous progress of science and technology and the continuous optimization of algorithms and data technology,the probability of overfitting will be minimized, interpretability will be improved, and the regulatory system will be improved, machine learning medical diagnosis technology will be more accurate, efficient and personalized, thus strengthening the construction of human health undertakings.

References

[1]Zhou Zhihua. machine learning [M]. Beijing: Tsinghua University Press, 2016:1-20.

[2]Dai Zhenmin, Sun Kuanhong, Chen Hao. Optimization algorithm of image object detection based on convolutional neural network. [A]. Yangzhou

[3]Zheng Huiwen. Application of machine learning algorithms in the diagnosis of medical diseases. Harbin (TP 181)

[4] Gao ge, Ma Shuai, Wang Xiaoying. Basic principles and application progress of computer-aided diagnosis in medical imaging diagnosis [A] 2016.12

[5]Zheng Ke Xin. Application of AI-based machine learning in medicine [A] 2022.9

[6]Ding Shihong. Machine learning-based medical-assisted diagnosis of breast cancer. TH865.2023.3eng Ke Xin. Application of AI-based machine learning in medicine [A] 2022.9