Machine Learning and Its Algorithms and Development

Ashish Mehra

Abstract: The basic classification of machine learning is look into in this article, which includes supervised learning, unsupervised learning, semi-supervised and reinforcement learning. It examines common machine learning methods such as the decision tree algorithm, random forest algorithm, artificial neural network algorithm, SVM algorithm and Boosting and Bagging algorithm. The goal is to increase people's awareness of machine learning and speed-up its popularization through the development of theoretical systems, further improvement of autonomous learning capabilities, the integration of multiple digital technologies, and the promotion of personalized custom services.

1. Introduction

Artificial intelligence has assist in new development opportunities alongside the rapid advancement of science and technology. Multidisciplinary theoretical knowledge, such as statistics and algorithm complexity, is incorporated into machine technology based on computer technology, which increases the functional features of artificial intelligence. It can provide direction for later machine learning development by conducting a realistic analysis of machine learning algorithms, consequently boosting the applicability of machine learning algorithms and giving greater convenience for the industry's economic development.

2. Basic Classification of Machine Learning

2.1. Supervised Learning

The supervised machine learning algorithms are those algorithms which needs external assistance. The input dataset is divided into train and test dataset. The train dataset has output variable which needs to be predicted or classified. All algorithms learn some kind of patterns from the training dataset and apply them to the test dataset for prediction or classification [1].

In the process of machine learning, supervised learning belongs to a relatively basic learning method. During the initial training of the machine, the machine relies on information technology to learn the needs of learning. In order to collect basic data information, we are supposed to gradually complete the required learning content in a supervised environment. Compared with other learning methods, supervised learning can fully stimulate the generalized learning potential of the machine itself. After completing the system learning, it can help people to solve some classification or regression problems, which is highly systematic. Currently, the classic learning methods commonly used include BN, SVN, KNN, etc.

2.2. Unsupervised Learning

Unsupervised learning is the inverse of supervised learning. Unsupervised learning indicates that the machine does not mark the material in a certain direction during the learning process, instead relying on the machine to complete the data analysis. In practice, the operating approach is to teach the machine the fundamental concepts and content, then give it enough freedom to perform a series of content learning tasks, such as learning concepts and content that are comparable to the fundamental principles, such as tree roots. The scope of machine learning content has grown as a result of the ongoing advancement of learning in phases. Deep belief networks and auto encoders are examples of unsupervised learning techniques currently available. Such circumstances are favourable for clustering problems to be solved, and they have a wide range of applications in the development of many industries.

2.3. Semi-supervised Learning

Semi–supervised learning algorithms is a technique which combines the power of both supervised and unsupervised learning. It can be fruit-full in those areas of machine learning and data mining where the unlabeled data is already present and getting the labeled data is a tedious process [2].

2.4. Reinforcement Learning

Reinforcement learning is a type of learning which makes decisions based on which actions to take such that the outcome is more positive. The learner has no knowledge which actions to take until it's been given a situation. The action which is taken by the learner may affect situations and their actions in the future. Reinforcement learning solely depends on two criteria: trial and error search and delayed outcome [3].

In addition to supervised learning and unsupervised learning, there are also application methods of reinforcement learning in machine learning. The data acquired in the prior period will be used in the specific application procedure. It forms a closed loop of data processing by organizing and processing the feedback information of a specific part. Overall, reinforcement learning is a form of learning strategy that uses statistics and dynamic learning to expand data collection. These strategies are mostly used to solve the challenge of robot control.

3. Analysis of Commonly Used Algorithms for Machine Learning

3.1. Decision Tree Algorithm

Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed on the basis of features of the given dataset. It is a graphical representation for getting all the possible solutions
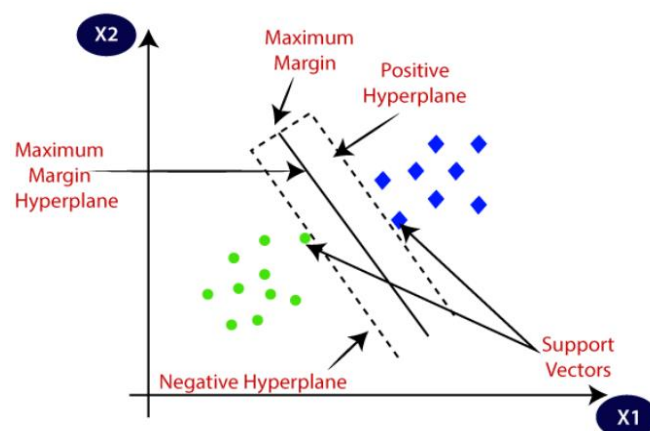
to a problem/decision based on given conditions. It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure. In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm. A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into sub-trees.

3.2. Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

3.3. SVM Algorithm

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyper-plane. SVM chooses the extreme points/vectors that help in creating the hyper-plane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyper-plane:

3.4. Boosting and Bagging Algorithm

Boosting algorithm as a new type of machine algorithm content, its biggest application advantage is that it can complete the accurate processing of data information and improve the accuracy of the final processing result. In practice, the function prediction system will be developed using the Boosting algorithm, and the system content will be continuously optimized using reinforcement learning mode, resulting in faster data processing. In the Boosting algorithm, AdaBoost is a reasonably simple application. At the same time, AdaBoost is a vital assurance for the Boosting algorithm's expansion. . The Bagging algorithm has a high similarity in the data processing process. In actual application, the difference is that the Bagging algorithm randomly selects the training set. The Bagging algorithm does not assess the weight content during the generation of the function model, therefore we must constantly modify the data model with the help of training to increase the accuracy of the data analysis findings.

4. Research on Machine Learning Development

4.1. Theoretical System Continues to Mature

The mechanical theory system will be further streamlined and its content branches and coverage will be enhanced in the future development process. The content of machine learning content was mostly applicable to some automation industries during the early formulation process, and the content of the overall theoretical system was not totally sound. The material of its theoretical theory is not applicable in several domains in practice. In response to such circumstances, the next stage of machine learning theory will be continuously improved, as will the degree of content refinement, providing favourable conditions for machine learning's following advancement.

4.2. Autonomous Learning Ability is Further Improved

At present, many enterprises in China have realized the development model of automation, and intelligence is the focus of the next stage of development. In the context of the rapid development of Internet technology, the autonomous learning ability of machines will be further strengthened. Whether it is supervised learning or unsupervised learning, the autonomy that machine learning can master will continue to increase. In the future learning process of the machine, the machine will perform targeted or extensive learning according to its own needs, which also reduces the economic cost of the enterprise to update the equipment structure, thereby laying a solid foundation for the stable development of the enterprise economy.

4.3. Integration of Multiple Digital Technologies

At this stage, relying on Internet technology has produced many branch technologies, such as Internet of Things technology, digital technology, cloud computing technology, etc. These technologies can provide many convenient conditions in the process of data calculation. Although these digital technologies are still in the initial stage of integration, with the rapid development of technology, the integration of digital technology is also constantly improving. Besides, in the future development process, these technologies will be combined with algorithms to form a new technology application system, thereby laying a foundation for the further improvement of data analysis speed.

4.4. Promotion of Personalized Customization Services

With the continuous improvement of socio-economic level, people's requirements for personalized applications are also constantly rising, which is also one of the important development directions of machine learning in the future. With the continuous improvement of the intelligent level of mechanical learning, different application modules can be set up according to the actual needs of users. After obtaining the user request message, the data module can filter out the corresponding information content and match the corresponding service content at the same time to meet the user's personalized needs and improve user service satisfaction.

5. Conclusion

In summary, machine learning is still in its infancy, and it mainly relies on supervised learning, and does not fully overcome weak artificial intelligence. Relevant personnel need to constantly improve the theoretical foundation and practice of machine learning. In the corresponding scientific field and the development of computer technology, we should provide a good environment for machine learning, and the development prospect of machine learning is very broad. In addition, it is also necessary to actively learn from the experiences and lessons of developed countries, set up machine algorithms suitable for the development of domestic enterprises, and provide technical support for the economic development of the industry.

References

[1] W. Richert, L. P. Coelho, "Building Machine Learning Systems with Python", Packt Publishing Ltd., ISBN 978-1-78216-140-0

[2] X. Zhu, A. B. Goldberg, "Introduction to Semi – Supervised Learning", Synthesis Lectures on Artificial Intelligence and Machine Learning, 2009, Vol. 3, No. 1, Pages 1-130

[3] R. S. Sutton, "Introduction: The Challenge of Reinforcement Learning", Machine Learning, 8, Page 225-227, Kluwer Academic Publishers, Boston, 1992