BDA5-2

1

Hello everyone, I am Haiying Che, from Institute of Data Science and knowledge Engineering

School of Computer Science, in Beijing Institute of Technology , in this session we discuss Data processing algorithm

2

Big data processing algorithm basically include data mining algorithms and Machine learning algorithms.

Machine learning algorithms include supervised learning, unsupervised learning semi-supervised learning and reinforcement learning.

Supervised learning is a method used to enable machines to classify objects, problems or situations based on related data fed into the machines.

Machines are fed with data such as characteristics, patterns, dimensions, color and height of objects, people or situations repetitively until the machines can perform accurate classifications.

Supervised learning is a popular technology or concept that is applied to real-life scenarios.

Supervised learning is used to provide product recommendations, segment customers based on customer data, diagnose disease based on previous symptoms and perform many other tasks.

Unsupervised learning is a method used to enable machines to classify both tangible有形的and intangible objects without providing the machines any prior information about the objects.

The things machines need to classify are varied, such as customer purchasing habits, behavioral patterns of bacteria and hacker attacks.

The main idea behind unsupervised learning is to expose the machines to large volumes of varied data and allow it to learn and infer from the data.

However, the machines must first be programmed to learn from data.

And in the scope of supervised learning there are regression (linear regression and polynomial), classification (logistic regression, SVM, ANN and decision trees), and deep learning (CNN, RNN).

And in the scope of unsupervised learning there are clustering (like K-means), dimensionality reduction (PCA), and anomaly detection.

For semi-supervised learning, it includes self-training and low-density separation models.

and reinforcement learning include dynamic programming and Monte Carlo methods.

3

Machine Learning is the scientific study of statistical models and algorithms that computer systems use to perform a task without explicit instructions.

Machine Learning is a comprehensive field that involves various functionalities of machine learning operations like clustering, classification and development of predictive models.

Basically, Machine Learning allows computers to learn without an explicit need for programming.

In general programming scenarios, you have to provide instructions to the computer for it to provide you with the output.

However, with the help of machine learning algorithms, you can train your computer to provide you with the output with the need to give instructions.

A machine learning algorithm is able to do so with the help of data. Using the data fed to the system, a machine learning algorithm is trained to provide output to the users.

Reinforcement Learning(RL) is a type of machine learning technique that enables an agent to learn in an interactive environment by trial and error using feedback from its own actions and experiences.

Though both supervised and reinforcement learning use mapping between input and output, unlike supervised learning where feedback provided to the agent is correct set of actions for performing a task, reinforcement learning uses rewards and punishment as signals for positive and negative behavior.

As compared to unsupervised learning, reinforcement learning is different in terms of goals. While the goal in unsupervised learning is to find similarities and differences between data points, in reinforcement learning the goal is to find a suitable action model that would maximize the total cumulative reward of the agent. The figure below represents the basic idea and elements involved in a reinforcement learning model.

4

Deep Learning is a recent field that occupies the much broader field of Machine Learning.

Deep Learning is most famous for its neural networks such as Recurrent Neural Networks RNN, Convolutional Neural Networks CNN, and Deep Belief Networks.

While other machine learning algorithms employ statistical analysis techniques for pattern recognition, Deep learning is modeled after the neurons of the human brain.

They are modeled after the structure and functioning of the human brain. In order to understand deep learning, we have to understand how the nervous system in the human body works. As we all know that our nervous system is built up of neurons. These neurons are able to grasp information that is transmitted to our body. These neurons have the ability to learn information over time. This principle of ‘learning’ is also utilized by artificial neural networks.

Any Deep neural network will consist of three types of layers: The Input Layer; The Hidden Layer and The Output Layer

What's the difference between the two?

Simply explained, both machine learning and deep learning mimic the way the human brain learns.

Its main difference is therefore the type of algorithms used in each case, although deep learning is more similar to human learning as it works with neurons.

Machine learning usually uses decision trees and deep learning neural networks, which are more evolved. In addition, both can learn in a supervised or unsupervised way.

5

from the table we can see , DL requires large data, ML can train on less data, DL can provide high accuracy but ML gives less accuracy; DL takes longer time to train, but ML takes less time to train,

L requires GPU to train properly and ML trains on CPU, DL can be tuned in various different ways, but ML has limited tuning capabilities.

6

Data mining algorithms can be categorized into 4 groups, classification, clustering, correlation Analysis and anomaly detection.

And the top 12 data mining algorithms is shown in the diagram, which include SVM, decision trees, neural network, ID3 algorithm, C4.5 algorithm K-nearest Neighbors, Naïve Bayes algorithm etc.

7

In this session we generally summarized the big data processing algorithms, which include the machine learning algorithms and data mining algorithms.

thank you for your attention, if you have any question, feel free to contact me.