

Make Python Learning Vivid and Efficient Through a Variety of Methods

Yutong Xie

Shandong Experimental High School, Jinan City, Shandong Province, 250001, 17805310012@139.com

Abstract: Python is one of the easy-to-learn languages and is easy to code and read, with high expressiveness. This article focuses on introducing basic content of learning Python, and discusses the ways to learn it efficiently with related cases, for the reference of learners.

Keywords: Python; Learning; Efficient Methods

Introduction Python is an object-oriented programming language released in the year of 1989. Python syntax is concise and clean, with large class library that be extended to other languages to complete module production. It can also be well embedded in C language and C++. These advantages make Python not only easy to read and code, but also elegant and concise, which may be favored by users. Therefore, it is important to explore efficient learning methods for applications of Python.

1. Basics about Python learning

1.1 Python programming language

Python programming language is relatively simple and programmer-friendly in terms of syntax. Therefore, learners are convenient to know which problems remain to be solved in each language when coding and reading. Moreover, Python is easy to understand, which is friendly for beginners to learn. Python is open-source which means it can be applied regardless varies systems and platforms, allowing free than ever network communication, and it also has different types of third-party libraries.^[1]

1.2 Machine learning

In terms of machine learning, while data is input into a system, corresponding answers can be obtained. With the help of data model and rules, it is applied to new data to generate the answer. Active learning is acquired through training instead of programming.

Advantages of machine learning are listed as following contents. Firstly, machine learning is more accurate, compared with human learning, that key points to solve problems can be found through data. When the amount of data increases gradually, the accuracy of data will become higher. Secondly, the active learning mode is characterized by automation, which means it will automatically complete learning a new mode and can be embedded into the automatic work flow. It makes the learning process more rapid because results will be obtained within a few milliseconds after data is input, showing a real-time response and having good effect. Thirdly, it can be customized. Most data-driven problems can be solved by machine learning. Learners are allowed to use their own data to complete the learning model construction and optimize it by using any standard. Fourthly, the use of machine learning can also effectively deal with the problem of data growth and the use of cloud computing to deal with large-scale data in the process of business

development.

1.3 Deep learning

Deep learning, which is similar to the brain neural structure, is a mathematical model belonging to network assists information processing. Convolutional neural network (CNN) is one of the typical image data processing models, in which there are two-dimensional convolution layers with spatial information such as height and width. Pictures are a matrix composed of pixels, and each pixel has three RGB channels, which enable continuously expansion of data volume, up to three times. If a programmer wants to convert pictures into data and input them into a computer, the data volume will be enormous, slowing down the running speed of the system. At this time, CNN can be used for cross-correlation operation instead of full connection to process pictures efficiently. In neural network, each neuron only needs to perceive local pixels instead of all pixels of the image to combine local information and obtain image representation information, which is quite comprehensive.

2. Case study of Python

2.1 Emotion analysis

2.1.1 Obtain data

This article chooses the analysis of stock data as a case through Python to carry out emotion analysis. All data is obtained from the network for the analysis of text emotion. Although there are application class interfaces in all kinds of social networks, there is no interface that can directly obtain the information of listed companies and a stock up to now. It is necessary to utilize the interface of application programs to obtain tweets first, and then filter out information related to stocks. Crawlers can also be coded to grab relevant information directly, with the judgment of keywords. However, the application of anti-crawler technology in social networks is constantly changing, which leads to the fact that the extraction of data features in this way is not completely reliable, and it is difficult to successfully apply to the production and trading environment.

2.2.2 Use word segmentation

Most titles are coherent sentences, and the use of word segmentation can obtain coherent sentences according to specific rules and recombine them into different sequences. For a simple sentence, it can be changed into different word combinations to form different results through word segmentation. At the same time, meaningless stop words or punctuation can be filtered out by word segmentation according to your requirements.^[2]

2.2.3 Text vectorization

After data acquisition and information filtering, feature extraction can be completed. The so-called text vectorization means explaining the semantic meaning of the text by means of numerical vectors. Generally, there are several ways to express the vectorization. First, the One-Hot method: the presented words all have their numerical values of 1, and the unexposed words all have their numerical values of 0 in the word sequence. Second, TF method: the numerical value of the presented words is equal to the frequency of the words appearing in the text, while the numerical values of the words not presented in the word sequence are all 0 in the word sequence. Third, TF-IDF method: the number of words presented in the word sequence is equal to the product of the frequency of appearing texts and the frequency of inverse documents, and the number of words not presented in the word sequence is equal to the presented in the word sequence is all 0. The third text vector method is selected below.

2.2.4 Classify emotions

According to the characteristics of text vectorization, it is regarded as an input feature. Then the prediction of headline emotion in stock review is carried out to judge whether it is positive emotion or negative emotion, and the classification of emotion is completed with the help of support vector machine.

2.2 Analysis of travel data

In the learning process, real data provided by Mobike is selected, and travel data are analyzed in different time and space dimensions.

2.2.1 Describe the original data

Original data set contains the following information: first the order number, second the vehicle ID, third the user ID, fourth the vehicle type, fifth the riding start date and time, sixth the riding start block position, and seventh the riding destination block position. Data from 0: 00 on May 10 to 24: 00 on May 23, 2017 are selected below, including 3,085,688 order data, 348,066 user data and 481,757 vehicle data.

2.2.2 Analyze time characteristics

According to the daily order numbers shown in Figure 1, it can be found that the number of rides generated every day had been decreasing since May 20th, which may be related to Mobike's sales strategy.



Figure 1. Chart of daily order numbers.

According to the information given in Figure 2, order volume generated in each period can be learned. The number of people who traveled by bike before 5: 00 am was relatively small, which could be ignored. The peak started from 7: 00 am to 9: 00 am and from 17: 00 pm to 19: 00 pm, which was obvious. In addition, there was also a small peak of riding during the lunch break.



Figure 2. Number of riding orders in each period.

2.2.3 Analyze the spatial characteristics

In figure 3, distribution characteristics of riding distance can be found. The average riding distance of all users is 815m. Among these users, 75% people could reach 949m, with 660m of median distance. From the above data, most users who chosen Mobike were riding at close range, so the analysis results are relatively reasonable. However, the farthest riding distance reached 44,900m, which is very likely to be an abnormal value. Therefore, value that is higher than 5km can be eliminated, and then the analysis shows that the distance that most users choose to ride is no more than 1000m.



Figure 3 User's cycling distance distribution

3. Conclusion

In summary, the above analysis indicates that by dint of its features, Python language can be widely used to solve life problems if one learns its basic knowledge. Therefore, learners are suggested to have a further study on Python, and improve the effectiveness of Python learning by applying and mastering its grammar to eventually solve a series of problems related to big data in practical cases.

References

- [1] Wang Z. Using various ways to make Python learning vivid and efficient (in Chinese). Course Education Research 2018; (1): 238-239.
- [2] Liao J. Research on the application of deep learning processing based on Python (in Chinese). Digital Space 2020;(3): 152.