PERFORMANCE ANALYSIS OF PROGRAMMING LANGUAGES USED IN MACHINE LEARNING MODELS

Fahad Molla, Chadel Dowridge, Herbert Furman, Aparicio Carranza

Computer Engineering Technology

19th Annual City Tech Poster Session

ABSTRACT

Machine learning (ML) is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. ML provides the user with insights and analysis to massive amounts of data by way of algorithms: computer code written to learn patterns in datasets and make predictions or inferences. The purpose of this study is to determine which of several programming languages is best suited to executing ML algorithms in specific environments. The programming languages C++, Java and Python are a vehicle to execute in different environments four environments: Windows, Linux Virtual machines (Parrot OS & Ubuntu) and Raspbian Linux distribution installed in a Raspberry PI. The metrics gathered include time to completion, memory utilized and CPU usage. Finally, we present to the reader observations made as to which language is superior per each environment.

INTRODUCTION

- The purpose of this project is to assess the performance of different programming languages when given machine learning tasks
- We will compare three languages: Java, C++ and Python
- We will make use of at least two IDEs: primarily Eclipse and Visual Studio
- We will use programs called “Profilers”, such as Visual VM, to gather data on running code
- We will compile the data gathered across different environments and machines and assemble them into a coherent performance review

LANGUAGES

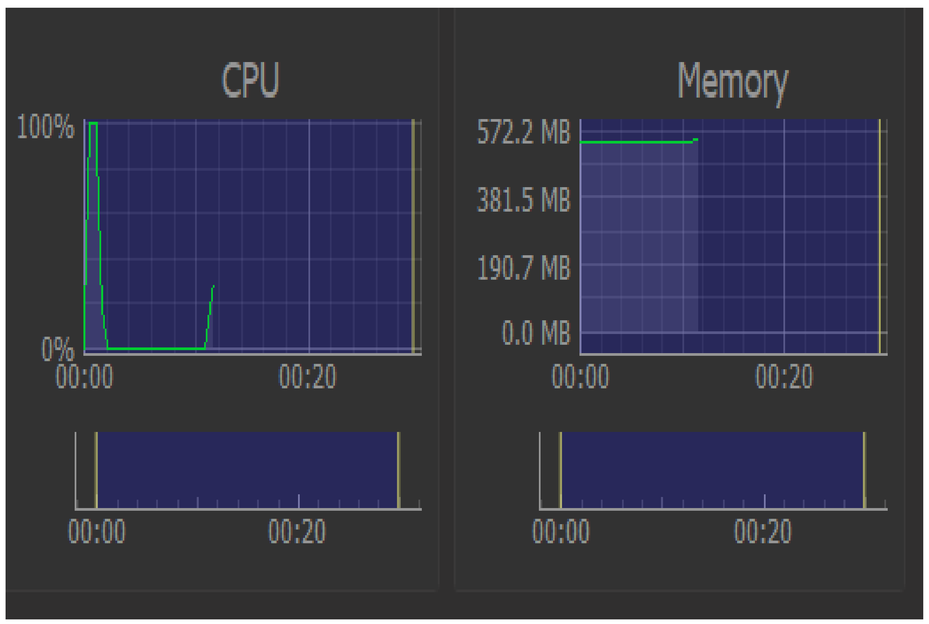
IDE

ENVIRONMENTS

METHOD – PROFILING

- Sniff passively all threads, CPU usage, memory usage for each language in accordance with its requisite profiling software
- VisualVM for Java
- PyVM Monitor for Python
- VisualStudio C++
- General Profilers

RESULTS



A screen capture of PyVMMonitor profiling a Python script for a K-Means Clustering algorithim

CONCLUSION

- C++ uses the least amount of memory
- Python is the slowest, but has the least CPU Usage
- Java is the quickest to create a model and display results
- Java may be best for tasks the require lots of data very quickly, like object avoidance
- Python may be best for tasks on microcomputers with limited CPU cores

REFERENCES

[1] Alpaydin, Ethem. Introduction to Machine Learning, MIT Press, 2014.

[2] Kaluza, Bostjan. Machine Learning in Java, Packt Publishing, 2016.

[3] Various. Introduction to Software Engineering, Wikibooks, 2021.