Science and data manipulation in Pharo

Polymath, Pharo-AI and DataFrame

Ferlicot-Delbecque Cyril | ESUG 2023
cyril@ferlicot.fr
Summary

- History
- What is new?
- Toward a new stage
History

Polymath

DataFrame

Pharo-AI
PolyMath

- Computation library for Pharo
- Similar to NumPy and SciPy in Python or SciRuby in Ruby
- Originally SciSmalltalk in Squeak
- Present in Pharo since a long time
### DataFrame

<table>
<thead>
<tr>
<th></th>
<th>sepal length (cm)</th>
<th>sepal width (cm)</th>
<th>petal length (cm)</th>
<th>petal width (cm)</th>
<th>species</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>5.0</td>
<td>3.4</td>
<td>1.6</td>
<td>0.4</td>
<td>setosa</td>
</tr>
<tr>
<td>28</td>
<td>5.2</td>
<td>3.4</td>
<td>1.5</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>29</td>
<td>5.2</td>
<td>3.4</td>
<td>1.4</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>30</td>
<td>4.7</td>
<td>3.4</td>
<td>1.6</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>31</td>
<td>4.8</td>
<td>3.1</td>
<td>1.6</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>32</td>
<td>5.4</td>
<td>3.4</td>
<td>1.5</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>33</td>
<td>5.2</td>
<td>4.1</td>
<td>1.5</td>
<td>0.1</td>
<td>setosa</td>
</tr>
<tr>
<td>34</td>
<td>5.5</td>
<td>4.2</td>
<td>1.4</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>35</td>
<td>4.9</td>
<td>3.1</td>
<td>1.5</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>36</td>
<td>5.0</td>
<td>3.2</td>
<td>1.2</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>37</td>
<td>5.5</td>
<td>3.5</td>
<td>1.3</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>38</td>
<td>4.9</td>
<td>3.6</td>
<td>1.4</td>
<td>0.1</td>
<td>setosa</td>
</tr>
<tr>
<td>39</td>
<td>4.4</td>
<td>3.0</td>
<td>1.3</td>
<td>0.2</td>
<td>setosa</td>
</tr>
<tr>
<td>40</td>
<td>5.1</td>
<td>3.4</td>
<td>1.5</td>
<td>0.2</td>
<td>setosa</td>
</tr>
</tbody>
</table>
DataFrame

- Table data structure
- Similar to DataFrame in Pandas, Julia, ...
- Heavily used in data science
- Created in 2017 during GSoC
Pharo-AI

• Created in 2020

• Implements classical machine learning algorithms (not deep learning)
  ○ K-Means, Linear Regression, N-Gram Model, …
Harmony of communities

Polymath

DataFrame

Pharo-AI
What is new?
Polymath: Modularization

- Rearchitecture
  - Extraction of data structures and random generators
  - Extraction of distributions in progress
- Cleaning of internal dependencies
Polymath

- Improvement of the CI robustness
- Align some conventions with Pharo-AI
- Divers cleanings and bug fixes
- Pharo 11 compatibility
Pharo-AI : Data manipulation

- Data partitioners : create tests sets
- Imputers : fill missing values
- Encoders : Standardize your datas
- Normalizer : Use common scales in your project
Pharo-AI

- Uniformization of projects
- Documentation
- Graph algorithms updates
- Divers speed up
- Cleaning and bug fixes in algos
DataFrame

- Speed up
- Better integration with other collections
- New visualizations based on DataFrames
- Integration with pharo-AI data preprocessing
DataFrame: GSoC 2023

- GSoC of Joshua Jose Dias Barreto
- Implementation of missing features
  - Better sorting
  - Data manipulation
  - Missing values management
  - ...

Google Summer of Code
DataFrame : GSoC 2023

Further improvements of DataFrame inspector
Toward a new stage
A push from the students

- DataFrame was started as a GSoC
- First AI algorithm were students projects

=> Data science interest more and more people
Agile Artificial Intelligence

Alexandre Bergel
RelationalAI, Switzerland

https://bergel.eu
We are answering to this call

- Engineers are now pushing those projects
- Projects are maintained
- Speed is becoming correct
Are you using scientific computing or data science?

Is the speed enough for you?

Are you encountering any problem?

Are you missing features for your projects?

Let us know ;)