

2016 年度 Ruby Association 開発助成メンター報告書

「Ruby を用いた初等統計解析の整備と構築」

村田賢太*

2017 年 3 月 31 日

概要

2016 年度 Ruby Association 開発助成に採択され実施されたプロジェクト「Ruby を用いた初等統計解析の整備と構築」について、メンターの立場から成果を評価して報告する。

1 プロジェクトの概要

1.1 プロジェクトの目的

本プロジェクトは芦田恵大、三軒家佑将、西田孝三の 3 名により「Ruby をデータサイエンスにおいて実用に足る言語にすること」を目指して実施された。

プロジェクトは上半期と下半期の 2 期に分けて進められた。

上半期では、「既存 gem ライブラリを組み合わせたデータ入力・解析・可視化の再現性を保証した実用例を解説したドキュメントの整備」を進めるに当たって、IRuby を利用するための環境構築インフラの整備と、データフレームのライブラリである daru の調査が実施された。

下半期では、上半期の成果に基づき下記の項目が実施された。

- IRuby を安定的に導入・実行するための改善
- daru の実装改善と機能強化
- numo-narray の daru の内部構造への適用
- 可視化ライブラリと daru の連携強化

1.2 成果の概要

本プロジェクトの成果をまとめると以下の通りである。

- IRuby を用いた初等統計解析のための環境構築インフラの整備
- 上記に関する日本語ドキュメントの整備
- 可視化ライブラリと daru との連携強化
- 再現性を保証したデータ解析の実用例を Jupyter notebook として作成
- 既存 gem が機能改善や機能追加を行える実装品質ではないことを明らかにした

そして、最後の成果に関連し、当初予定していたが未達となってしまった目標が存在する。それは次の 2 点である。

- daru の実装改善と機能改善
- numo-narray の daru の内部構造への適用

これらの未達目標に代わり、numo-narray を内部構造として利用したデータフレーム・ライブラリの新規実装が試みられた。

1.3 メンターとしての関係性

メンターである筆者は、プロジェクト実施者の連絡用 Slack チャンネルに招待され、そこで必要に応じて意見を伝えたり質問をしたり、逆に実施者からの質問に答えるなどの対応を随時行なった。

* muraken@gmail.com

2 プロジェクトに対する評価

2.1 メンターから見たプロジェクトの位置づけ

本プロジェクトは、Ruby のために開発された独自のデータ解析環境の整備として解釈でき、現時点で Ruby が強みを発揮できていない分野に対して Ruby の将来性を高めるための取り組みとして重要である。

メンターである筆者は、本プロジェクトとは異なるアプローチとして、Python などの他言語へのブリッジを整備することで Ruby をデータサイエンスに適用可能とするプロジェクトを進めている [1]。

本プロジェクトと筆者のプロジェクトは、互いに異なる長所と短所を持っており、相互に補完し合う関係である。そのため、両者のアプローチが同時に進捗することは Ruby のデータサイエンス分野における将来性を考えると非常に望ましいことである。

2.2 特筆すべきプロジェクト成果

本プロジェクトの特筆すべき成果は次の 2 点である。

- 既存 gem が機能改善や機能追加を行える実装品質ではない問題を明らかにした
- 新規にデータフレームライブラリの開発を開始した

1 点目は、この問題を解決しなければ Ruby のための独自の仕組みの開発を進めることは困難となる、という性質を持っている。そのため、本プロジェクトは Ruby におけるデータサイエンスのためのエコシステムを開発していく上で非常に重要な知見を身をもって明らかにした。

2 点目に関しては、データフレームライブラリについて、よりシンプルな作りが可能であることを立証するための試みである。これはまだ開発が始まったばかりの段階であり、ライブラリそのものを評価できる段階ではない。

2.3 今後の開発に対する期待

本プロジェクトの今後の開発について次の点に期待したい。

1 点目はインフラ整備である。本プロジェクトで整備した IRuby に関するミドルウェアの整備は、Ruby や OS のバージョンが上がることで、そのままでは対応できなくなる可能性が高く、継続的なメンテナンスが必要となるだろう。そこで、継続するためのスキームの構築を期待したい。

2 点目は実用例のドキュメンテーションに関することである。現時点では著作権に関わる問題があり、成果を公開できていない。本プロジェクトでの成果を各 gem ライブラリのドキュメントの形に昇華させるなどの工夫を施すことで、成果が無駄にならないようにしていただきたい。

最後に再現性のある実用例に関すること。本プロジェクトでは初等統計解析を対象としたものに限定されているが、これをさらに応用に踏み込んだ実用例にまで広げていくことを期待したい。そうすると、業界トレンドの変化に追従していく必要が生じ、インフラ整備と同様に継続的なアップデートが必要になるだろう。これについても、継続できるスキームの構築までを含めて期待したい。

3 まとめ

本レポートでは、2016 年度 Ruby Association 開発助成に採択され実施された「Ruby を用いた初等統計解析の整備と構築」プロジェクトについて、メンターの立場からプロジェクトの成果を評価し、今後の開発に対する期待を述べた。

参考文献

- [1] 村田賢太。データサイエンスにおける Ruby の現在の位置づけと可能性。しまねソフト研究開発センター研究報告 (2016 年 12 月 5 日) <http://bit.ly/2okzW7X>