

数理・情報のフロンティア
2021 年度採択研究者

| |
|------------------|
| 2021 年度 年次報告書 |
|------------------|

中丸 智貴

東京大学 大学院総合文化研究科
助教

試行錯誤を効率化するための Jupyter Notebook 拡張

§ 1. 研究成果の概要

2021年10月から開始した研究であり、査読付き論文や公開済みのソフトウェアなどの成果はまだないものの、研究は申請時から大きく進展し、申請当初よりも具体的かつ学術的・社会的価値の高い技術の提案と開発を進めることができています。

本研究で目指しているのは、データサイエンスやAI構築などに取り組む際に避けることができないプログラムの試行錯誤を効率化し、最終的にはデータを活用して社会や学問の発展を加速することである。データの分析方法や予測モデルの種類、それらのハイパーパラメータを事前に決定することは難しく、プログラムの修正と結果比較を繰り返すという試行錯誤を行い、実験的に最善策を見つけるのが一般的である。しかしながら、現状そのような試行錯誤が行われているとは言い難い。愚直にプログラムの修正と実行を繰り返し、試行間で共通する部分を試行のたびに再計算していることは頻繁にある。本研究が目指すのは、そのような再計算を行う無駄を削減するための新しいソフトウェア技術及びそれを実装した基盤ソフトウェアの開発である。

上記の目標を達成するため、申請当初は抽象度の高いレイヤー(具体的にはPython言語レベル)でのキャッシュ/メモ化の仕組みの開発を試みていた。しかしながらプログラミング実装技術の研究者との議論を重ねた結果、そのような方法では実用に耐えうるような時間/空間コストで試行錯誤の効率化が行えない(効率化のためにソフトウェア的に行う操作のオーバーヘッドが大きく、再計算を減らすことができてでも作業時間は増大してしまう)可能性が非常に高いことが明らかになった。

そこでこの議論を踏まえ、より抽象度の低いレイヤー(具体的にはC言語レベル)でプログラムが引き起こすデータの変更を追跡し、効率的な試行錯誤を実現する手法を採用することになった。より具体的には、言語処理系が行うメモリ等の資源管理に関連するコードを改変し、高速かつ空間効率も良い形で不必要な再計算が起こらないようにする仕組みを採用すると言うことである。本研究では特に、そのような言語処理系の改変を自動で行うような技術の開発を行う。自動で改変を行う技術が構築できれば、元の言語処理系自体のコードに大幅な更新がある場合、今回対象としているPython以外の言語処理系について取り組む場合にも、素早く試行錯誤の効率化を実現できることが見込まれる。このような言語処理系の自動改変技術は、申請時に提案したPythonに特化した手法よりも応用先も多く、学術的/社会的価値は高い方法になっていると考えられる。分野内の研究者との議論により研究が進んだ結果である「抽象度の低いレイヤーを取り扱う手法」は既に大局的な設計は済んでいる状態であり、現在その着想を実際に動作するソフトウェアとして実装を進めている段階である。

ここまでで述べた研究の核となる技術の開発以外にも本研究の進展は挙げられる。本研究で開発した技術は最終的にJupyter NotebookというデータサイエンスやAI構築の分野で人気のあるプログラミング環境と統合し、実際に試行錯誤を効率化できるかなど、実証試験を行うことを目指している。実証試験を行うには具体的なユーザインターフェイスも重要になるが、この部分については既に基本的な実装を完了させ、操作中の画面を撮影したデモ動画まで作成している。このようなデモ動画を作成することで、プログラミング言語の実装技術を専門としない研究者等に対し、具体的にどのようなことをどう解決しようとしているのかを視覚的に示すことができるため、さまざまな場で役立つ。