

[事業報告]

講義におけるGoogle Colaboratory導入の試み

浅野 比

山陽小野田市立山口東京理科大学 共通教育センター

An Attempt to Introduce Google Colaboratory in a Class

Hitoshi ASANO

Center for Liberal Arts and Sciences, Sanyo-Onoda City University

要約

山陽小野田市立山口東京理科大学(本学)は、2021年度に文部科学省「数理・データサイエンス・AI教育プログラム認定制度(リテラシーレベル)」及び「数理・データサイエンス・AI教育プログラム認定制度(リテラシーレベル)プラス」に選定された。また、2023年4月には、工学部に数理情報科学科(仮称)を新設予定であり、データサイエンス教育の充実化を図っている。データサイエンスでは統計や機械学習を用いた分析が必要であり、プログラミングのスキルは必須である。本講義で導入したGoogle Colaboratoryは、Pythonをブラウザから実行できるアプリケーションで、無料で利用可能、環境構築不要、機械学習やディープラーニングの環境の構築が可能、CPU及びGPU環境を利用可能などの多くの利点を有する。現在プログラミングの講義がない応用化学科の学生に対してプログラミングへの抵抗を少しでも和らげ、興味を喚起するために、1コマと少ない回数ではあるが、担当講義である分析化学においてGoogle Colaboratoryによるプログラミングの回を取り入れた。本稿ではGoogle Colaboratoryを使用した図表の作成内容及び講義終了後のアンケート結果について報告する。

Key words: Google Colaboratory, programming, data science

キーワード: Google Colaboratory, プログラミング, データサイエンス

1. 緒言

山陽小野田市立山口東京理科大学(本学)は、2021年度に文部科学省「数理・データサイエンス・AI教育プログラム認定制度(リテラシーレベル)」及び「数理・データサイエンス・AI教育プログラム認定制度(リテラシーレベル)プラス」に選定された¹⁾。また、2023年4月には、工学部に数情報科学科(仮称)を新設予定であり²⁾、データサイエンス教育の充実化を図っている。近年の情報社会において、大量に存在するデータを適切に収集・管理し、その分析結果を社会問題の解決や課題発見などに有効に役立てることのできる人材を育成することは高等教育機関の重要な役割である。

2020年には小学校でプログラミング教育が必修化され³⁾、中学校⁴⁾及び高等学校でもそれぞれ2021年及び2022年からプログラミングに関する学習が拡充されている⁵⁾。「数理・データサイエンス・人工知能(AI)」が、これからの「読み・書き・そろばん」となり⁶⁾、大学や高専でも文系・理系問わず、データサイエンスのリテラシーレベルの引き上げが求められている。データサイエンスにおいて統計や機械学習、深層学習を用いた分析が必要である場合が多く、そのためにはプログラミングのスキルが必須である。したがってプログラミング教育は極めて重要であり、プログラミングの講義がない応用化学科の学生に対してプログラミングへの抵抗を少しでも和らげ、興味を喚起するために、1コマだけではあるが、担当講義である分析化学においてプログラミングに触れる機会を設けた。近年、化学分野にもAI、機械学習や深層学習が導入され、構造と物性の関係のモデリング⁷⁾、合成計画⁸⁾、分子設計⁹⁾、反応条件探索¹⁰⁾、創薬¹¹⁾において大きな進歩をもたらしている。また、分析化学の分野でも、分子構造から溶解度の予測¹²⁾、スペクトルデータの分類¹³⁾、NMRデータのノイズ除去¹⁴⁾、ウイルスの識別¹⁵⁾などに機械学習が用いられている。

プログラミングはプログラミング言語といわれるように、文字通り言語であるため、利用頻度を増やすことで理解が深まり、定着するが、時間的な制約もあり、本講義では体験するレベルにとどまっている。本稿ではGoogle Colaboratory¹⁶⁾を使用した表及び図作成を行った内容及びアンケート結果について報告する。

2. Google Colaboratory

プログラミング言語のPythonは比較的学習が容易であり、初学者向けである一方、Scikit-learn, TensorFlow, PyTorchなどの機械学習やディープラーニングでも利用で

きるライブラリが豊富であり、基礎から応用・実用まで幅広く利用されている。そのPythonをブラウザから実行できるアプリケーションが、Googleが開発したGoogle Colaboratoryである。Google Colaboratoryの特徴として、無料で利用可能、環境構築不要、機械学習やディープラーニングの環境の構築が可能、CPU及びGPU環境を利用可能などがあげられる。一方、利用にはGoogleアカウントの作成が必要、オフラインでは利用できない、時間制限がある(90分で切断状態、12時間で強制的に切断状態となる)、データは記録されているなどがあるが、私用フリーメールとしてGoogleアカウントを所持している学生は多い。また、時間制限も90分1コマの講義時間では無視でき、記録されて困るような重要データも扱っていないため、デメリットはほとんどない。講習会や実験実習形式ではなく、1コマ90分と時間も限られているため、ソフトの(やや煩雑な)インストールが不要で、すぐに使用可能なGoogle Colaboratoryは、プログラミングに触れるアプリケーションとしては最適であると考えられる。プログラミング言語名をWeb検索し、ヒット件数をランキングしている(人気の指標)TIOBE Index¹⁷⁾によると、Pythonは1位である(Fig.1, 2022年11月現在)。

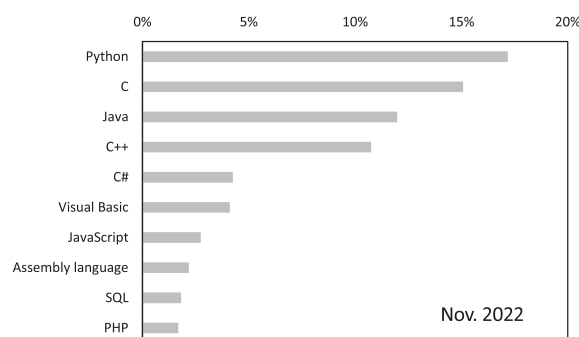


Fig.1 プログラミング言語のランキング

3. 実施内容

3.1 エチレンジアミン四酢酸のpHに対する各化学種のモル分率変化

分析化学は、応用化学科2年生向け専門科目、選択の講義である。講義概要としては、溶液内の現象を平衡(酸-塩基、錯形成、溶解、酸化還元)の概念から取扱い、理解するものである。その中で、錯形成において重要なキレート剤である、エチレンジアミン四酢酸(EDTA)を取り上げ、pHによってその化学形態が変化することを平衡論から理論的に学習した。理論式からpHにおける各化学種のモル分率を算出し、グラフ化することでpHと化学種のモル分率を

視覚的にとらえるため、Google Colaboratoryを使用して、表及びグラフを作成することを主な内容とした。機械学習や深層学習を扱うプログラミングにおいて、データの読み込み、表形式の表示及び結果のグラフ化は極めて基本的で、必要かつ不可欠な操作であり、1コマ分という時間的な制約もあることからこれらの作業に絞って実施することとした。

3.2 Google Colaboratoryのインストール

フリーメールであるGmailのアカウントについては、プログラミングの講義の回までに取得することを初回のガイダンスやMoodle上で告知していたこともあり、新しくアカウントを作成しなければいけない学生は数人程度であった。メール(Gmail)やオンライン学習(Google Classroom)などにGoogleのサービスを導入している大学等もあるが、その際はインストールに制限がある場合がある。Anacondaは、オフラインでPythonのプログラミングが可能であるが、初学者にとってインストールはやや難しいと思われる。また、時間的な制約もあるため、気軽にPythonに触れるという面では比較的インストールが簡便であるGoogle Colaboratoryは非常に優位である。しかしながら、予想はしていたが、インストールがスムーズに行えない学生も5人程度いたため、それぞれ個別に対応することに苦慮した。プログラミングの講義をスムーズに行うために、事前にGoogle Colaboratoryのインストールを実施しておくことが必要であった。インストールに関しては、web上に動画をはじめとしてわかりやすい手順の説明があり、学生1人でも事前にできたことである。

3.3 プログラミング

実際に実行したコード及びグラフをFig.2に示す。初めに表及びグラフを扱うためのライブラリである「pandas」及び「matplotlib.pyplot」をそれぞれインポートした。その後、データもとである、csvファイルを読み込ませるためのドライブのマウント、ファイルの読み込みを行った。今回は与えられたコードを入力するという形式で行ったが、単に入力するだけではタイピングの練習で終わってしまうため、コードの意味、グラフに関する事項(軸タイトル、プロットの色、凡例)について、どこを変更するとどう変わるかについて、時間のゆるす限り説明し、体験してもらった。

プロトン濃度[H⁺]、pH及び各化学種のモル分率をあらかじめエクセルで算出しておき、csvファイルとしたものをGoogle Driveに保存し、保存したドライブのマウント、csvデータの読み込みのコードを入力させ、実行させた。当然ながら、インポートライブラリ名のスペルミス、「.(ドット)」

が、「(カンマ)」(またはその逆)、全角で入力などの理由で、実行時にエラーを生じる学生がいた。エラーが生じると、エラー内容が表示されるため、その内容から自分で修正を施すようにと指示を出したが、英語であるため内容を把握することがやや難しく感じている印象を受けた。したがって、エラーに関しては個別に対応する必要があり、TAなどの補助があれば、もう少しスムーズに進行できた。

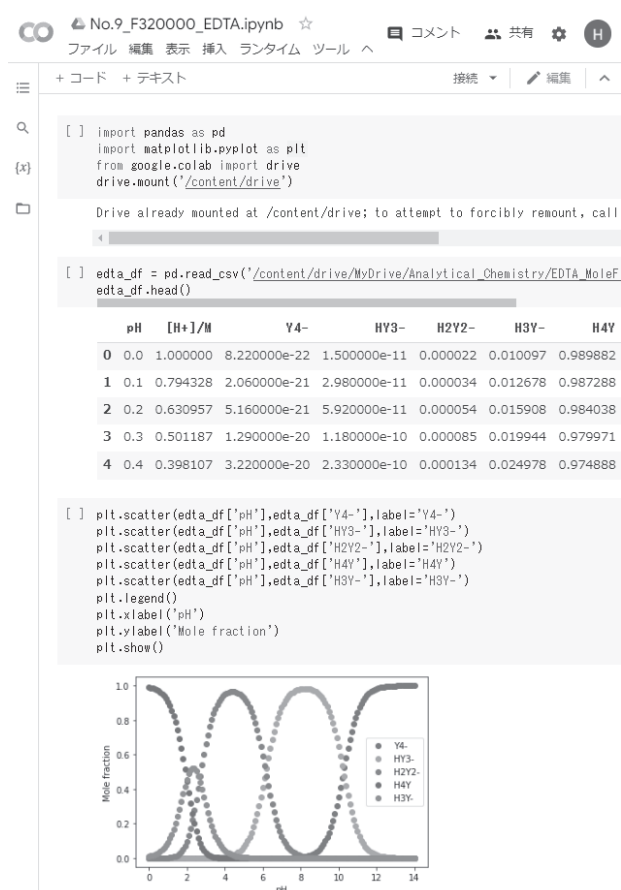


Fig.2 Google Colaboratoryによるプログラミング画面のスクリーンショット

4. アンケート結果

プログラミングが終了したのから、以下の7項目についてアンケートを実施し、受講生44名から回答を得た。

- ① Google Colaboratoryを知っていましたか?
- ② プログラミングをやったことがありますか?
- ③ ②で「はい」と回答した方に質問です。やったことのあるプログラミング言語は何ですか? (Python・C言語・JavaScript・C++・C#・Ruby・VB・PHP・SQL・MATLAB・スクラッチ・その他から複数選択可)
- ④ プログラミングの難易度はどうでしたか?
- ⑤ プログラミングスキルの習得は必要だと思いますか?

⑥ 化学でも機械学習(ML)やディープラーニング(DL)などの知識が必要だと思いますか？

⑦ 今後プログラミングを勉強したいと思いますか？

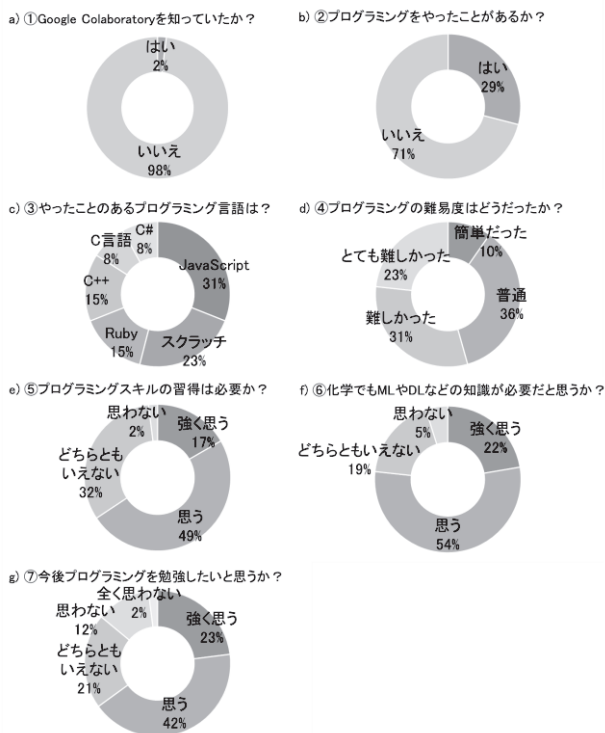


Fig.3 アンケート集計結果

Fig.3(a)より, Google Colaboratoryについて知らなかった学生の割合が98%となり, 認知度は極めて低いことがわかった. このことから, 16回の講義のうちのたった1回ではあるが, 実施してよかったと感じた. 一方, Fig.3(b)及び(c)に示すように, プログラミングを行ったことがある学生は3割で, JavaScript:31%, スクラッチ:23%と半数を占め, 次に, C++及びRuby:15%, C#:8%, C言語:8%の順であった. 本学でも5年ほど前に小中学生向けにスクラッチを使用したプログラミング講座を実施していたが, アイコンを移動させるだけの簡単なスクラッチは, 初中等教育課程で触れる機会があったものと考えられる. その他のプログラミング言語の選択肢としては, Python, Java, VB, PHP, SQL及びMATLABを用意したが, いずれもやったことのある学生はおらず, その点についてもPython(Google Colaboratory)を実施した意義があったと感じている. なお, やったことのあるプログラミング言語は複数回答を含んでいる. 一方, Fig.3(b)に示すように, プログラミング経験者は29%であったが, 15行程度の短いコードで, 簡単だったと解答する学生は10%(普通も含めて46%)で, 難しいと感じる学生は54%と約半数であった. プログラミングを行ったことがない, 雰囲気もわからない学生にとって,

単純なコードであっても難しいと感じることはごく自然だと思われる. プログラミング経験者からすると, プログラミングすると, どのような感じで実行されるかを体験しているため, 初めて使用したプログラミング言語でも, Pythonのシンプルさもあって敷居を下げる事が出来たのではないかと考えられる. 単純な作業でも, プログラミングに少しでも触れておくことは重要であると考えられる. プログラミング経験者29%で, 難易度の感想が簡単~普通で46%であり, おそらくこの範囲に経験者が含まれていると考えられるが, 未経験者でも17%の学生は, 難易度を普通と感じていることが推測される. 今回の講義では, 1コマ内で行える内容としてデータ読み込み, 表作成, グラフ化を目的とし, 難易度についてはあまり考えていなかったが, 内容の設定としてはアンケート結果から概ね妥当であったと考えられる. 次のプログラミング等の必要性に関するアンケート項目に関しては, Fig.3(e)及び(f)に示すように, プログラミングスキルの習得については66%の学生が必要であると考えており, 76%の学生が化学でも機械学習, ディープラーニングの知識は必要だと考えている. これらから, 7割程度の学生がデータサイエンスへの興味, 必要性・重要性については認識があるものと推察される. 一方, どちらともいえないが, Fig.3(e)及び(f)でそれぞれ32%及び19%おり, 将来それらに関する職業に就くとも限らない, 化学系だからプログラミングにかかわる部分はあまりないと考えて, どちらともいえないという選択肢になっているのかもしれない. 文部科学省はプログラミングに携わる職業を目指す, 目指さないにかかわらず, 初等教育の段階からプログラミング的思考を育むことを目的として, これからの時代において共通に求められる力であると位置づけし, プログラミング教育を実施している. 最後の項目(Fig.3(g))の今後プログラミングを勉強したいと思った学生は半数以上の65%となり, プログラミング学習への意欲は持っているようである.

5. 結論

今回はGoogle Colaboratoryを講義に導入し, その実施内容, アンケート結果について報告した. プログラミングは分野問わず極めて重要なスキルとなっているが, 半期の講義中の1コマの講義では, 触れた程度で, 当然ながらプログラミングスキルを習得するレベルまで持っていくことは困難である. 今回実行したコードは, 15行程度の簡単なもので, 具体的にはドライブのマウント, csvファイルの読み込み, 表の表示及びグラフ化で, プログラミングの恩恵を感じられる内容にはなっていない. しかしながらこの短時間の

講義を通じて、興味を持った学生は自分で学習を進めたり、今後プログラミングの機会がある場合は、アレルギー反応を起こすことなく、スムーズに学習するためのきっかけになれば幸いである。

参考文献

- 1) 山陽小野田市立山口東京理科大学, 数理・データサイエンス・AI教育プログラム,
<https://www.socu.ac.jp/departments/ai.html> (accessed 2022-11-08)
- 2) 山陽小野田市立山口東京理科大学, 数理情報科学科,
<https://www.socu.ac.jp/departments/informatics-data-science.html> (accessed 2022-11-08).
- 3) 小田理代, 後藤義雄, 星 千枝, 永田衣代, 青木 譲, 赤堀侃司: 各教科等横断的なプログラミング教育の実践による小学校教師の変容に関する考察 - Technological Pedagogical Content Knowledge (TPACK)の形成の観点から-, *STEM教育*, 2, 3-13, (2020).
- 4) 野口祥太, 小倉 康: 中学校段階における体系的なプログラミング教育カリキュラムの研究—理科授業へのプログラミング導入を核として—, *日本科学教育学会研究会研究報告*, 35, 33-38, (2020).
- 5) 兼宗 進: 初等・中等教育の新教育課程で変わる情報教育, *システム/制御/情報*, 62, 248-253, (2018).
- 6) 鈴木静男, 田中康平, 大沼 巧, 玉置慎吾, 大庭勝久, 酒井基至, 竹口昌之, 白井秀範, 芳野恭士: 衛生データプラットフォームを用いた遠隔方式によるデータサイエンス教育, *工学教育*, 69(4), 80-85, (2021).
- 7) Mater, A. C., Coote, M. L.: Deep Learning in Chemistry, *J. Chem. Inf. Model*, 59, 2545-2559, (2019).
- 8) Coley, C. W., Thomas, D. A., Lummiss, J. A. M., Jaworski, J. N., Breen, C. P., Schultz, V., Hart, T., Fishman, J. S., Rogers, L., Gao, H., Hicklin, R. W., Plehiers, P. P., Byington, J., Piotti, J. S., Green, W. H., Hart, A. J., Jamison, T. F., Jensen, K. F.: A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science*, 365, eaax1566, (2019).
- 9) Mater, A. C., Coote, M. L.: Deep Learning in Chemistry, *J. Chem. Inf. Model*, 59, 2545-2559, (2019).
- 10) Kondo, M., Wathsala, H.D.P., Sako, M., Hanatani, Y., Ishikawa, K., Hara, S., Takaai, T., Washio, T., Takizawa, S., Sasai, S.: Exploration of flow reaction conditions using machine-learning for enantioselective organocatalyzed Rauhut-Currier and [3+2] annulation sequence, *Chem. Commun.*, 56, 1259, (2020).
- 11) Sanchez-Lengeling, B., Aspuru-Guzik, A.: Inverse molecular design using machine learning: Generative models for matter engineering, *Science*, 361, 360-365, (2018).
- 12) Delaney, J. S.: ESOL: Estimating Aqueous Solubility Directly from Molecular Structure, *J. Chem. Inf. Comput. Sci.*, 44(3), 1000-1005, (2004).
- 13) Ofner, J., Kamilli, K. A., Eitenberger, E., Friedbacher, G., Lendl, B., Held, A., Lohninger, H.: Chemometric Analysis of Multisensor Hyperspectral Images of Precipitated Atmospheric Particulate Matter, *Anal. Chem.*, 87, 9413-9420, (2015).
- 14) Kusaka, Y., Hasegawa T., Kaji, H.: Noise Reduction in Solid-State NMR Spectra Using Principal Component Analysis, *J. Phys. Chem. A*, 123, 10333-10338, (2019).
- 15) Taniguchi, M., Minami, S., Ono, C., Hamajima, R., Morimura, A., Hamaguchi, S., Akeda, Y., Kanai, Y., Kobayashi, T., Kamitani, W., Terada, Y., Suzuki, K., Hatori, N., Yamagishi, Y., Washizu, N., Takei, H., Sakamoto, O., Naono, N., Tatematsu, K., Washio, T., Matsuura, Y., Tomono K.: Combining machine learning and nanopore construction creates an artificial intelligence nanopore for coronavirus detection, *Nature Commun.*, 12(1), 1-8, (2021).
- 16) Google Research, Welcom to Google Colaboratory, <https://colab.research.google.com/> (accessed 2022-11-08).
- 17) TIOBE, TIOBE Index, <https://www.tiobe.com/tiobe-index/> (accessed 2022-11-08).