

ドリトルでのデータ処理機能とグラフ描画機能の開発

小林 史弥^{1,a)} 白井 詩沙香^{2,b)} 山澤 昭彦^{3,c)} 兼宗 進^{1,d)}

概要: 高等学校における共通教科「情報」は、2018年度改訂の新学習指導要領において共通必修科目「情報Ⅰ」、選択履修科目「情報Ⅱ」の2科目に再編される。「情報Ⅱ」では「情報とデータサイエンス」として、データ分析に関する内容も扱われることとなり、扱いやすい学習環境や学習教材等の開発・研究が求められている。本研究では、高等学校において容易にデータサイエンスを学ぶことができる学習環境を提供することを目的に、教育用プログラミング言語「ドリトル」のデータ処理機能およびグラフ機能を開発する。本稿では、ドリトルに基本的なデータ処理機能および「度数分布表」や「統計グラフ」等のグラフ描画機能を実装したので報告する。

キーワード: データサイエンス, 学習環境, ドリトル

Data Processing and Graph Drawing Function in Dolittle

KOBAYASHI FUMIYA^{1,a)} SHIRAI SHIZUKA^{2,b)} YAMAZAWA AKIHIKO^{3,c)} KANEMUNE SUSUMU^{1,d)}

Abstract: New course of study for high schools will be published in fiscal 2018. Under new course of study, Informatics will be realigned into two subjects which are the common compulsory course “Information I” and the elective course “Information II”. In the chapter of “Information and Data Science” in “Information II”, students will learn about data analysis. To conduct effective teaching for Data Science will require the development of learning tools that are easier for students to handle. In this study, we developed new function in the educational programming language “Dolittle” for learning Data Science. In this paper, we report the implementation of basic data processing functions and graph drawing functions.

Keywords: Data Science, Learning Tool, Dolittle

1. はじめに

高等学校における共通教科「情報」は、2018年度改訂の新学習指導要領において共通必修科目「情報Ⅰ」、選択履修科目「情報Ⅱ」の2科目に再編される。特に、情報Ⅱで

は、「情報とデータサイエンス」というデータ分析に関する項目が入ることで注目されている [1]。現行の学習指導要領においても、問題解決のための手段として簡単なデータ分析が取り扱われており、教科書では主に表計算ソフトウェアを利用したデータ処理、グラフ描画が紹介されている [2]。しかし、表計算ソフトウェアの利用は便利な反面、関数の挿入やグラフの描画をするためには、リボンのタブ切替やセルの範囲選択といった複数の操作が必要となり、教員は操作に慣れていない生徒に指示を出すのが難しく、生徒自身も操作手順に気を取られ、データ処理の流れを意識しづらいという課題がある。

本研究では高等学校において容易にデータサイエンスを学ぶことができる学習環境を提供することを目的に、教育用プログラミング言語「ドリトル」のデータ処理機能およ

¹ 大阪電気通信大学
Osaka Electro Communication-University, Neyagawa, Osaka
572-8530, Japan

² 武庫川女子大学
Mukogawa Women's University, Nishinomiya, Hyogo 663-
8137, Japan

³ イーテキスト研究所
eTextLab, Hachioji, Tokyo 192-0352, Japan

a) mm17a006@oecu.jp

b) shirai@mukogawa-u.ac.jp

c) a.yamazawa@gmail.com

d) kanemune@gmail.com

びグラフ機能を開発する。ドリトルを用いることで、操作の指示はプログラムを提示するだけで行え、データ処理の流れをそのままプログラムに書き起こすことができる。本稿では、ドリトルに基本的なデータ処理機能およびグラフ描画機能を実装したことを報告する。

2. 先行研究

2.1 データサイエンス学習環境

現在、高等学校の共通教科「情報」では、「社会と情報」、「情報の科学」の両教科において、問題解決のためのコンピュータ活用が取り扱われている [3]。本単元において、問題解決の手法の一つとして簡単なデータ処理を学ぶ。取り扱われる問題は、地球温暖化のような社会問題や学園祭の出し物を決めるといった生徒に身近なものである [4], [5]。例えば、地球温暖化を題材とした場合は、CO₂ 排出の最大値、最小値を求めて排出量の幅を把握したり、平均値を出したり、それを大きく上回るものを探して排出量が多くなる原因を探ったりする。これらのデータ処理には、表計算ソフトウェア (Excel) が用いられることが多い。Excel は操作をよく理解している者が、個人で利用する際にはデータの集計や加工、グラフ描画などを行う際に有効なソフトウェアである。しかしながら、操作をよく知らない生徒が授業の中で利用するには問題がある。第一に、操作の説明が容易ではないという点である。たとえば、Excel で値の平均を求めるときには、リボンから数式パレットを開き、AVERAGE 関数を挿入し、対象のデータをドラッグして選択する。もしくは、セルに「=AVERAGE(A1:A10)」と入力するといった方法がある。どのセルに数式を入れるのか、数式パレットはどこにあるのか、指定範囲はどこからどこまでかということをお口頭で伝えると、生徒には理解が難しい。第二に、処理の順番などを生徒が考えることが難しいという点である。生徒の思考が、平均を出すための Excel の操作手順を考えるといった「Excel の個々の操作」になってしまう。そのため、平均を計算し、次に分散を求めるといった「データ処理の流れ」という思考になりにくい。

Excel 以外のデータ分析のためのツールとして R 言語 [6] や SPSS Statics [7] などがある。R 言語は、コマンドラインから操作する CUI の統計解析用言語である。RStudio [8] といった開発環境がフリーで配布されており、当該環境を使えば GUI での利用が可能である。高度な解析・分析機能を備えており、大学の講義などの教育現場から、実務まで様々な場面で利用されている。SPSS Statics は IBM が提供する有料の統計解析ソフトウェアで R 言語と同様に高度な解析・分析を行うことができ、実務で利用されることが多い。両ソフトウェアともに有用なソフトウェアであるが、高等学校での利用を考えたとき、商業用のソフトウェアは導入の面において敷居が高く、ソフトウェアのインストールが必要なものは、導入が容易でないことが予想される。

```
1 かめた=タートル! 作る。
2 L字=かめた! 50 歩く
3     90 左回り
4     100 歩く
5     図形を作る。
```

図 1 ドリトル言語でのプログラム例
Fig. 1 Sample program of Dolittle

2.2 ドリトル言語

ドリトル言語 (以下、ドリトル) は教育用に設計されたオブジェクト指向プログラミング言語である。日本語でコードが記述でき、初中等教育現場で扱いやすいようオンラインプログラミング環境 Bit Arrow から利用できるオンライン版、Java によるローカル版ともに特別なインストール作業なしに利用できる。

ドリトルは「オブジェクトに命令を送る」プログラミングモデルであり、「結果のリターン値のオブジェクトに続けて次の命令を送る」というスタイルを利用して、シェルのパイプライン処理のようなプログラムが書きやすい。例えば、図 1 のような画面上に「L 字」を描くプログラムでは、1 行目で生成した「かめた」というタートルオブジェクトに対し、2~5 行目で「歩く」、「左回り」、「歩く」、「図形を作る」という 4 つの命令を続けて実行している。

このように、ドリトルで記述されたプログラムは、実行される処理の流れを把握しやすいという特徴をもつ。

こうした特徴から、ドリトルをデータサイエンス学習に利用することで、データ処理の流れをそのままプログラムに書き起こすことができ、生徒がデータ処理の手順を理解しやすくなると期待される。また、教員が操作手順を伝える場合にもプログラムを提示するだけで良く、指導しやすくなると考えられる。

現在、ドリトルには数値計算や図形描画、ファイルの読み込みなどの機能は既にあるため、従来のドリトルでも数値計算や図形描画と言った処理は可能である。しかし、統計・データ解析、統計グラフ用のメソッドがあるわけではないため、実行するためにはプログラムを一から記述する必要がある。

そこで本研究では、ドリトルでデータの分析や抽出、グラフの描画といった操作を簡単に行える機能を開発する。

3. ドリトルのデータサイエンス機能

3.1 必要な機能の検討

開発する機能を検討するために、高等学校の授業で行われているデータ分析の流れを図 2 に示す。例えば、前節で挙げた地球温暖化を題材とした授業において CO₂ の排出量を減らす方法を考える場合、まずはじめにゴミの焼却における素材ごとの CO₂ の排出量のデータを入力する (図 2 の (1))。次に、データを整理してデータを扱いやすくする

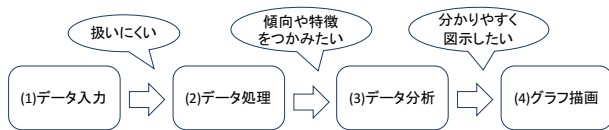


図 2 データ分析の流れ

Fig. 2 Flow of data analysis

ために排出量を大きい順や小さい順に並べ替えるといった処理を施す(図2の(2)).最後に、図2の(3)に示すように、データの傾向や特徴を掴むために、データの分析を行う。例えば、平均値を出すことでCO₂排出量の相場が明らかになり、平均値を大きく上回る素材のゴミを減らせば効果的にCO₂の排出量が削減できることがわかる。該当する素材のゴミの廃棄量を半分に減らした際の総排出量を求め、減らす前の総排出量と比較する場合は、差分を視覚化するためにグラフ化を行う(図2の(4)).他の題材についても、図2と同様の流れでデータ分析を行っていくことから、(1)~(4)の各処理で必要となる機能を実装していく。

3.2 機能の実装

前節で検討した処理を行うために実装した機能を表1に示す。まず、データ入力については、csv形式のファイルを読み込み、表形式のデータとして取り込む。このとき、csvファイルの1行目のデータをフィールド名とし、読み込んだデータを格納したオブジェクトをテーブルオブジェクトとする。読み込むデータは統計表のような集計済みのデータでも、データベースのような集計前のデータでも良い。

データ処理では、データの並べ替えに加え、選択・射影・結合といったデータベース操作言語と同様の操作ができるようにした。テーブルオブジェクトに対して条件式を与え、それに合致するデータを抽出することで実現する。「選択」の場合は条件式を引数にとり、条件に当てはまるデータを含む行だけを取り出し、オブジェクトに格納する。「射影」の場合はフィールド名を引数にとり、合致する列のデータだけをオブジェクトに格納する。また、データの並び替えはフィールド名を引数にとり、合致する列のデータをソートしオブジェクトに格納する。

データ分析では、分析を行うための様々な統計関数が必要となる。データ分析の命令は、フィールド名を引数にとり、合致する列の値をもとに計算しオブジェクトに格納するようにした。

グラフ描画は、フィールド名を引数にとり、合致する列の値をもとにグラフの描画を行う。

そのほかに、未集計のデータを分析するために表生成機能も実装した。フィールド名を引数にとり、合致する列のデータをもとに表を生成する。

表 1 用意した命令

Table 1 Function list

処理	命令	動作
データ入力	ファイルから作る	csv ファイルを読み込む
データ処理	選択	条件に合致する行を取り出す
	射影	列を取り出す
	結合	2つのデータをくっつける
	大きい順	データを大きい順に並べる
	小さい順	データを小さい順に並べる
データ分析	合計	合計を求める
	平均	平均を求める
	中央値	中央値を求める
	最大値	最大値を求める
	最小値	最小値を求める
	第1四分位数	第1四分位数を求める
	第3四分位数	第3四分位数を求める
	度数	度数を求める
	標準偏差	標準偏差を求める
	分散	分散を求める
クロス集計	クロス集計を行う	
グラフ描画	円グラフ	円グラフを描画する
	棒グラフ	棒グラフを描画する
	折れ線グラフ	折れ線グラフを描画する
	散布図	散布図を描画する
	箱ひげグラフ	箱ひげグラフを描画する
表の作成	度数分布表	度数分布表を作る

3.3 機能の使用例

実装した機能の使用例として、e-Statの「平成27年度青少年のインターネット利用環境実態調査」の「機器の利用状況」[10]の統計表を利用したデータ分析を紹介する。ここでは、上記の統計表を加工し都道府県ごとの機器利用の状況を示す統計表を作った一部を図3に示す。

大阪府と和歌山県を対象に、インターネット利用時の「ノートパソコン」利用者数を比較したプログラムを図4に示す。まず、データ入力操作として、1行目でcsv形式のファイル読み込みを行っている。次に、利用した統計表のうち、分析で使用する「ノートパソコン」のフィールドだけを取り出すために、2行目で「射影」機能を使っている。続いて、3行目では大阪府と和歌山県のレコードだけを取り出すためにOR演算を行い、「選択」機能を使ってレコードを抽出した。続いて、4行目では取り出したデータを表示している。最後に、5行目では取り出したデータをもとに、棒グラフを描画している。描画したグラフを図5に示す。

このように、プログラムで実行する処理の命令が、データ処理の流れにそのまま対応しているため、ユーザはプログラムを見返すことで、どのような手順でデータの分析を行ったかを理解することができると期待される。

次に、成績データをもとに箱ひげ図を書く例を紹介する。この例では、データをもとに「最大値」、「最小値」、「第1四

	A	G	H	I	J
1	都道府県	携帯電話	機能限定電話	ノートパソコン	デスクトップパソコン
25	三重県	7.7	3.8	7.7	7.7
26	滋賀県	-	11.4	20.5	4.5
27	京都府	8.9	8.9	24.4	6.7
28	大阪府	11.2	8.1	24.2	13.9
29	兵庫県	14.5	7.7	25.6	8.5
30	奈良県	16.4	10.9	36.4	14.5
31	和歌山県	12.5	3.1	9.4	15.6

図 3 分析するデータの例
Fig. 3 Sample data

- 1 データ=フィールド! "01.csv" ファイルから作る。
- 2 結果=データ! "都道府県" "ノートパソコン" 射影
- 3 「どなか! (都道府県 == "大阪府") (都道府県 == "和歌山県") 本当」選択。
- 4 結果! 表示。
- 5 結果! "都道府県" "ノートパソコン" 棒グラフ。

図 4 棒グラフを表示するプログラム例
Fig. 4 Sample program of bar graph

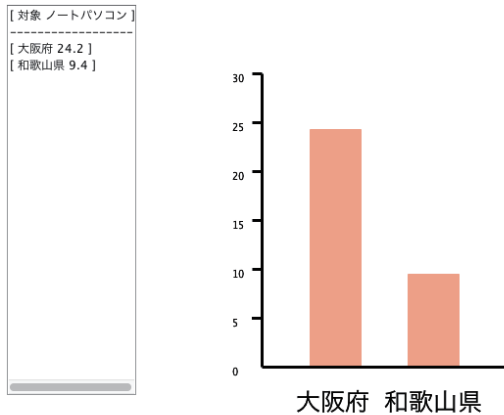


図 5 棒グラフを表示するプログラムの実行例
Fig. 5 Result of bar graph sample

- 1 成績=テーブル! "成績.txt" ファイルから作る。
- 2 成績! 表示。
- 3 成績! "組" "国" 箱ひげグラフ。

図 6 箱ひげグラフを表示するプログラム例
Fig. 6 Sample program of box plot

分位数, 「第3四分位数」, 「中央値」を内部的に計算し, グラフ化している。成績データは「番号」, 「組」, 「国」, 「数」, 「社」, 「理」, 「英」の7つのフィールドがある。図6にプログラム例を示す。

最初に, 1行目では成績データの読み込みを行っている。次に, 2行目では成績データを表示している。最後に, 3行目では「組」ごとの「国」の点数に関する箱ひげグラフを描画している。描画したグラフを図7に示す。

4. まとめと今後の展望

本研究では, 高等学校において容易にデータサイエンス

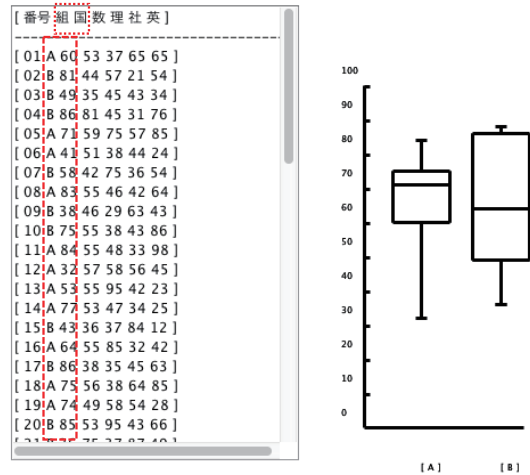


図 7 箱ひげグラフを表示するプログラムの実行例
Fig. 7 Result of box plot sample

を学ぶことができる学習環境を提供することを目的に, 教育用プログラミング言語「ドリトル」のデータ処理機能およびグラフ機能を開発した。本稿では, 基本的なデータ処理およびグラフ描画機能を実装し, 現在の高等学校の共通教科「情報」で行われているデータ分析を行えることを確認した。今後は, 次期学習指導要領から始まる「情報II」におけるデータサイエンス学習のための学習環境として利用できるようさらなる機能の追加を行っていきたい。

謝辞 本研究は JSPS 科研費 17K00989 の助成を受けたものです。

参考文献

- [1] 鹿野利春: 次期学習指導要領を見据えて これからの情報活用能力 (オンライン), 入手先 (<http://www.sky-school-ict.net/shidooryo/170106/>) (参照 2017-05-19).
- [2] 岡山敏雄, 山極隆: 最新情報の科学, 実教出版 (2015).
- [3] 文部科学省: 高等学校学習指導要領解説情報編 (2012).
- [4] 水越敏行, 村井純, 生田孝次: 見て分かる社会と情報, 日本文教出版 (2014).
- [5] 岡山敏雄, 山極隆: 高等学校社会と情報, 実教出版 (2014).
- [6] The R Foundation: The R Project for Statistical Computing (オンライン), 入手先 (<https://www.r-project.org/>) (参照 2017-05-19).
- [7] IBM: IBM SPSS Statics Japan (オンライン), 入手先 (<http://www-01.ibm.com/software/jp/marketplace/spss/>) (参照 2017-05-19).
- [8] RStudio: RStudio - Open source and enterprise-ready professional software for R (オンライン), 入手先 (<https://www.rstudio.com/>) (参照 2017-05-19).
- [9] 大阪電気通信大学, 東京農工大学, 明星大学: オンラインプログラミング環境 ビットアロー (Bit Arrow) (オンライン), 入手先 (<http://bitarrow.eplang.jp/index.php?FrontPage>) (参照 2017-05-19).
- [10] 総務省 統計局: 機器の利用状況 (オンライン), 入手先 (<https://www.e-stat.go.jp/SG1/estat/GL02020101.do?method=csvDownload&fileId=000007664116&releaseCount=1>) (参照 2017-05-19).