

ドリトルを用いたデータ活用授業の提案

小林史弥[†] 白井詩沙香^{††} 兼宗進[†]

[†]大阪電気通信大学 ^{††}大阪大学

1 はじめに

2018年に公示された高等学校共通教科「情報」の新学習指導要領では、データの活用およびデータサイエンスが扱われる [1]。従来はデータの整理から分析まで表計算ソフトウェアのみが利用されることが多かったが、情報システムにおけるデータの利用を考えた際に、プログラミングとの繋がりをイメージしづらいことなどが予想される。そこで、我々は初学者を対象としたデータ活用学習支援を目的とし、教育用プログラミング言語ドリトル [2] のデータ処理およびデータの可視化機能を備えたデータ分析ライブラリを開発してきた [3]。本稿では、本ライブラリを用いたデータ活用授業を設計したので、報告する。

2 ドリトルのデータ分析ライブラリ

2.1 データ分析ライブラリの概要

ドリトルは教育用に設計されたオブジェクト指向のプログラミング言語であり、日本語のプログラムを記述することができる。また、オブジェクトに命令を投げ、その結果を返すという方式であるため、UNIXのパイプラインのようなプログラムを書きやすい。例えば、図1に示すサンプルデータから通学手段が「電車」の生徒の平均通学距離を算出するには、図2のようなプログラムを記述する。データに対して命令を投げかけて、得られた結果に新たな命令を投げるモデルであるため、一連のデータ処理の流れが追やすい。そのため、データを絞り込んで必要な範囲のデータに統計処理を行うというデータ活用のプロセスを扱いやすいと考えられる [3]。また、命令名が日本語表記であるため、各プロセスで行った処理を理解しやすいという特徴をもつ。

図3に図1のプログラムを実行したドリトルの画面

Data Analysis Education Program using Dolittle
Fumiya KOBAYASHI[†], Shizuka SHIRAI^{††},
Susumu KANEMUNE[†]

[†]Osaka Electro-Communication University 572-8530,
Neyagawa, Japan

^{††}Osaka University 567-0047, Toyonaka, Japan

生徒番号	通学手段	住所	週あたりの読書時間	通学距離	学年
s28145	電車	生駒市	1	34.3	1
s28146	自転車	西宮市	5	2	1
s28147	徒歩	西宮市	2	0.3	1
s28148	電車	草津市	2	76.5	1
s28149	電車	神戸市	1	15.9	1
s28150	徒歩	西宮市	0.5	1	1

図 1: サンプルデータ (ba_sch.tsv)

```
1 データ=テーブル!"data/ba_sch.tsv" ファイル
  から作る。
2 電車データ=データ!"通学手段=="電車"」選択
  表示。
3 電車データ!"通学距離" 平均値 表示。
```

図 2: 電車通学の生徒の平均通学距離を算出する例

を示す。左側がプログラム入力画面であり、右側が実行画面である。実行画面では、命令の実行結果を実行前と見比べることができるため、実行した命令に対する結果がどの様になったかを確認しながら学習することができる。

2.2 機能

本ライブラリでは、テーブルオブジェクトというオブジェクトにデータを配列としてもたせており、ユーザはテーブルオブジェクトに対して様々な命令を実行し、データの分析や可視化を行うことができる。テーブルオブジェクトには以下の4つの機能がある。

データ入力機能 TSV や CSV の形式のファイルや、プログラム中で生成した配列からデータをテーブルオブジェクトに取り込める。読み込むファイルは図1のように1列目がフィールド名で、2列目以降がデータの形式とした。

データ操作機能 データベースオンライン学習支援システムである sAccess [4] との授業連携を想定し、選択、射影、結合といったリレーショナルデータベースの操作を行える。図1のデータから電車通学の生徒を抽出する場合には図2のプログラムとなる。

データ処理機能 平均値や最頻値、標準偏差の算出、度数の算出やクロス集計のような基本的な統計処理が行える。それぞれの命令は、空値を省いて計算を行い、計算結果を持ったテーブルオブジェクトを返す。



図 3: ドリトルの画面例

表 1: 提案する授業の構成

時限	概要	主な活動内容
1	ドリトルの導入と質的データの集計	度数の算出, クロス集計, 選択などの基本操作
2	量的データの要約	平均値, 中央値, 最頻値の算出, ヒストグラムの描画
3	適切なグラフの選択	各種グラフの描画
4	相関	相関係数の算出, 散布図の描画
5	データ分析課題	これまでの内容を踏まえてコンビニの売り上げデータを分析

データの可視化機能 棒グラフ, 積み上げ棒グラフ, ヒストグラム, 折れ線グラフ, 円グラフ, 帯グラフ, 散布図, 箱ひげ図等を描画できる。

3 データ活用授業の設計

本ライブラリを用いた授業を紹介する。本授業はドリトルのデータ分析ライブラリに同梱されている図1に示した「西宮市の高校に通う生徒」に関するデータを用いる。授業の構成を表1に示す。授業は5回の授業で構成される。

1回目では, データの種類について簡単に説明し, 質的データを扱う。質的データの集計手段として, 度数の算出とクロス集計のプログラムを作成する演習を行う。また, 「選択」命令によるデータの抽出の仕方も扱う。同時にドリトルの文法に慣れる必要があるので, プログラムの意味, 文法を丁寧に解説しながら進める。

2回目では, 量的データを扱う。平均値, 中央値, 最頻値をそれぞれ求め, 図4のようにヒストグラムと照らし合わせることで適切な代表値を決定する演習を行う。データの範囲を3年生に絞り込んだ際の代表値を決定するなどの課題を与える。

3回目では, 各種グラフを扱う。それぞれのグラフについて, 特徴を説明する。そして, 「一年生の生徒数を住所別に可視化するには何グラフが最適か?」といった練習問題をいくつか与え, 生徒に自力でプログラムを書くことに挑戦してもらう。

4回目では, 相関を扱う。この回では, 相関係数や散布図を描画する演習を行うが, それぞれの数値, グラフの解釈の仕方を解説することに重きをおく。

5回目では, これまでの内容を踏まえ, コンビニの売

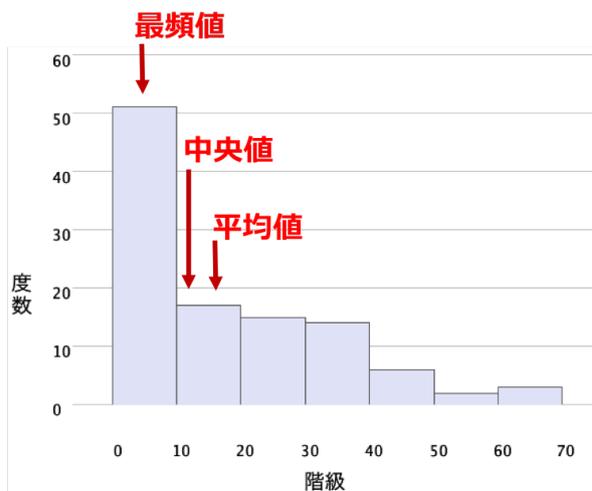


図 4: データの分布と各指標

り上げデータを与えて分析をしてもらう。客層や, 年代ごとの売れ筋などを分析し, コンビニの経営を改善するためのアイデアを考えてもらう。

4 まとめ

本稿では, 教育用プログラミング言語「ドリトル」のデータ処理および可視化機能を備えたデータ分析ライブラリによるデータ活用授業を紹介した。今後は, 本提案授業を実践したい。

参考文献

- [1] 文部科学省:高等学校学習指導要領解説 情報編, <http://www.mext.go.jp/component/a_menu/education/micro_detail/_icsFiles/afieldfile/2018/07/13/1407073_11.pdf> (参照 2019-01-15)
- [2] 兼宗進, 久野靖: ドリトルで学ぶプログラミング [第2版], イーテキスト研究所 (2011).
- [3] 小林史弥, 本多佑希, 白井詩沙香, 兼宗進: オンライン版ドリトルを用いたデータ分析学習環境の開発. 情報教育シンポジウム論文集. Vol.2018, No.16, pp.112-117(2018).
- [4] 長瀧寛之, 中野由章, 野辺緑, 兼宗進: データベース操作の学習が可能なオンラインの学習教材の提案. 情報処理学会論文誌. Vol.55, No.1, pp.2-15 (2014).