

解説ページに対する評価を用いた R 言語学習の不理解点の推定

Estimation of Understanding Failure in Learning R from Reputation for Explanation Pages

大江高悠[†]
Takahisa Oe

島川博光[‡]
Hiromitsu Shimakawa

1. はじめに

近年、コンピュータのハードとソフトの進化により、データ解析のブームがおきている。企業では、特に医療やビジネスの分野でデータ解析や解析結果活用のために統計解析の知識を持つ人材の需要が増えている。

R 言語は、統計解析のために作られた言語である。R 言語には統計解析を深くサポートしており、環境に依存せず手軽に扱えるという利点がある。R 言語を扱える人材が求められているが、現状 R 言語を使える人材が不足している。その背景の一つとして、教員に R 言語を教えるノウハウが蓄積していないことが挙げられる。

C 言語や Java などの汎用言語と違い、R 言語を扱うためには、統計学の知識が必要である。しかし R 言語を扱うために必要な統計学の知識は膨大である。現状の R 言語を教える参考書や参考サイトでは、その膨大な統計学の知識をカバーできていない。そこで、本研究では、R 言語の学習者に足りていない知識を推定する手法を提案する。学習者に足りていない知識をカバーできるよう R 言語教育を強化し、R 言語を扱える人材を増やす必要がある。

2. R 言語教育の現状

2.1 プログラミング教育の既存研究

プログラミング教育の重要性が高まっている現在、プログラミング教育支援のための研究が多くなされている。柏原らは、プログラム理解支援のために、プログラム空欄補充問題を自動生成する手法を提案した [1][2]。杉浦らは、アルゴリズム構成能力を育成するために、手作業でアルゴリズムを実行・理解させるという方法を提案した [3]。斐品らは、図を用いたアルゴリズム学習支援システムを用いることで、学習の効率が上がることを示した [4]。兼宗らは、初中等教育での情報教育のためのオブジェクト指向プログラミング言語であるドリトルを開発した [5]。西田らは、処理の流れなどのプログラミングの基礎を習得するための学習環境を開発した [6]。

これらの研究では、アルゴリズムなどの処理の流れを組む力が必要である C 言語や Java などの汎用言語を対象としている。しかし、R 言語を扱うための概念として、処理の流れ以外に、統計学の知識が必要である。

2.2 R 言語の記法と概念

R 言語を扱うためには、記法への知識と、概念(統計学)への知識が必要である。汎用言語では、処理の流れが大切であり、コードを読むことで理解が深まることがある。しかし、R 言語では、概念への知識が必要であり、コードを読むだけでは概念への知識は身につかない。

学習者の多くは、R 言語の記法が理解できていないと知っているが、実際は概念が理解できていない場合が多

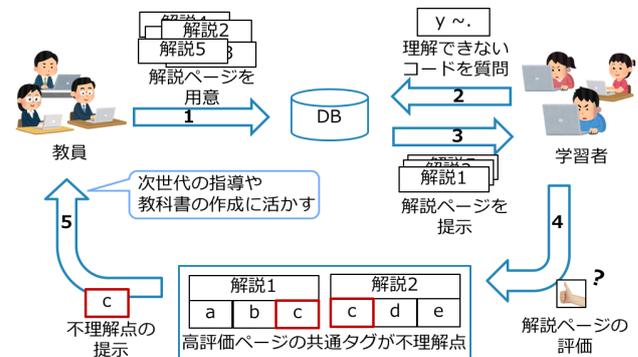


図 1: 不理解点の推定手法

い。例えば、「y ~ .」という記法が分からないときに、学生は記法の意味が理解できていないと考えることが多い。しかし、実際には、「~」がモデルの従うことを示し、「.」が入力データの全変数を説明変数とすることを意味する。よって、学生はこれらの概念が理解できていないことになる。

2.3 学習者の不理解点

本研究では、統計学についての深い知識を持たない、R 言語の初学者を学習者と呼ぶ。学習者の不理解点は、概念への不理解であることが多いが、多くの学習者は、それを自覚せず、記法への不理解だと考えている。本研究で不理解点は、学習者が理解できていない R 言語を扱うために必要な知識を指す。

教員が学習者の不理解点を考慮した指導ができていないことが、R 言語を使える人材が不足している原因のひとつとして挙げられる。

本論文では、学習者の不理解点を明らかにし、教員の指導を支援するシステムを考える。そのために、本論文では、教員に学習者の不理解点を提示する手法を提案する。

3. 学生の不理解点の推定手法

本研究では、学習者の不理解点を推定する手法について提案する。手法の概要を図 1 に示す。

本手法は以下の 3 つのフェーズに分けられる。

1. 教員が解説ページを用意するフェーズ
2. 学習者が理解できないコードを質問し、与えられた解説ページの評価をするフェーズ
3. 高評価ページの共通タグを不理解点として抽出するフェーズ

本手法は、学習者が理解できないと考える記法(コード)から、実際に理解できていない概念を教員に提示することを目的とする。

[†]立命館大学大学院情報理工学専攻

[‡]立命館大学情報理工学部

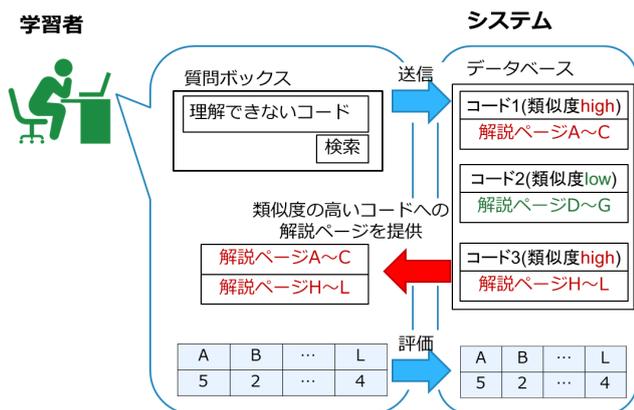


図 2: 解説ページの評価の方法

3.1 教員による解説ページの用意

教員は、理解されづらい R 言語のコードを収集する。理解されづらいコードは、R 言語初学者に R 言語の参考書などを提示し、理解することが難しかったコードを示してもらい、それをを用いる。また、理解されづらいコードについて、文字の量などの制限は設けない。

理解されづらいコードに対して、教員はいくつかの解説ページを用意する。解説ページは、教員が学習者の疑問を解決すると想定して用意した、Web 上の既存のページを指す。解説ページには、記法について説明したページや、概念について説明したページが含まれる。また、ひとつの記法に対して、関係する概念は複数あると考えられるため、ひとつの理解されづらいコードに対しての解説ページは複数になると考えられる。

例えば、「`ksvm(y ~ ., data = train)`」という理解されづらいコードに対して、教員は「サポートベクターマシンについて」、「`ksvm` 関数の使い方」、「formula の書き方」などの解説ページを用意する。

理解されづらいコードと、それに対する解説ページのペアを、システム上のデータベースに保存する。

3.2 学習者による解説ページの評価

学習者が理解できないコードを質問し、与えられた解説ページの評価をするフェーズについて、概要を図 2 に示す。

学習者は、理解できないコードを質問ボックスに入れる。理解できないコードは、学習者が R 言語の参考書などをを用いて学習している際に、理解ができなかったコードを示す。質問ボックスは、テキストボックスと送信ボタンで構成され、PC の画面上に表示させる。テキストボックスに理解できないコードをペーストし、送信ボタンを押すことで、システムに理解できないコードを送信する。

システムは、

- 質問ボックスで学習者から受け取ったコード
- 3.1 で教員が収集した理解されづらいコード

の間の類似度を算出する。算出した類似度が一定以上のコードに対する解説ページを、システム上のデータベースから学習者に提供する。

学習者は与えられた解説ページそれぞれに対して 5 段階での評価により点数を付ける。学習者の疑問が解消された場合や、新しい知識が得られた場合に解説ページに対して、学習者は高い点数を付ける。つまり、高い点数が与えられた解説ページは、学習者のもともとの不理解点を含むことになる。

3.3 不理解点の推定

与えられた点数の平均点が高い解説ページの、タグを取得する。本手法でタグは、はてなブックマーク [7] の機能を用いて、解説ページに対してはてなブックマークの他のユーザがつけたタグを利用する。

取得されたタグの中で、数が多かったものを、学習者の不理解点として教員に提示する。また、そのタグの付けられている解説ページ群の評価の平均点も、不理解の度合いとして教員に提示する。

教員は与えられた不理解点を解消できるように、次世代の指導や、教科書の作成に活かすことができる。

例えば、3.1 の `ksvm()` 関数の例で、不理解点として、「カーネル」が推定されたとする。その際、教員は `ksvm()` 関数やサポートベクターマシンを教える際に、「カーネル」についての知識と一緒に指導することで、指導をより良くしていくことができる。

4. おわりに

本論文では、R 言語の学習者の解説ページへの評価から不理解点を推定する手法を提案した。今後は提案手法の有効性を、実験により検証する。

参考文献

- [1] 柏原昭博, 久米井邦貴, 梅野浩司, 豊田順一: プログラム空欄補充問題の作成とその評価, 人工知能学会論文誌, 16 巻 4 号 C, pp.384-391 (2001) .
- [2] Akihiro. Kashihara, Atsuhiko. Terai, Jun'ich. Toyota: Making fill-in-blank programproblems for learning algorithm, ICCO'99, pp.776-783 (2001) .
- [3] 杉浦学, 松澤芳昭, 岡田健, 大岩元: アルゴリズム構築能力育成の導入教育: 実作業による概念理解に基づくアルゴリズム構築体験とその効果, 情報処理学会論文誌, 49 巻 10 号, pp.3409-3427 (2008) .
- [4] 斐品正照, 徳岡健一, 河村一樹: 構造化チャートを用いたアルゴリズム学習支援システム, 情報処理学会論文誌, 45 巻 10 号, pp.2454-2467 (2004) .
- [5] 兼宗進, 中谷多哉子, 御手洗理英, 福井眞吾, 久野靖: 初等教育におけるオブジェクト指向プログラミングの実践と評価, 情報処理学会論文誌, 44 巻 13 号, pp58-71 (2003) .
- [6] 西田知博, 原田章, 中村亮太, 宮本友介, 松浦敏雄: 初学者用プログラミング学習環境 PEN の実装と評価, 情報処理学会論文誌, 48 巻 8 号, pp.2736-2747 (2007) .
- [7] 「はてなブックマーク」, < <http://b.hatena.ne.jp/> > 2018 年 6 月 22 日アクセス .