

1Y-05

Git を使用した不適切な学習行動を取る学生の発見と手法

成田裕貴 広瀬啓雄
公立諏訪東京理科大学

1. はじめに

プログラミングの授業の課題の多くは、指定されたプログラムを作成し提出するものである。作成したプログラムは一人ひとり違うものであるはずだが、中には変数名やインデントを変えただけのものや他の学生のものを全てコピーしたようなものも存在する。提出されたものだけでは適切な成績をつけることが困難であり、正当に評価されるべき学生の期待を満たせないケースも存在する。

そのため、本研究ではプログラミング学習者の学習行動を明らかにし、適切な学習行動を取っていない学生を早期にピックアップする手法を確立することを目的とする。

本研究の意義は不適切な学習行動を取る学生の発見方法にある。GIGA スクール構想が既に実施されているが、適切な指導者がいない等の理由で導入に踏み込めない自治体も存在する。本研究ではプログラミングを知らなくてもデータ分析の結果を理解できるようにグラフ化しているため、指導方法確立の一要素になると考えている。

2. 実験

2.1 環境

- Python 3.10.0
- Docker version 20.10.10, build b485636
- Git version 2.30.2
- 編集ツール：visual studio code 1.63.0

2.2 実験方法

2.2.1 対象者及び期間

- 対象：広瀬研究室所属 3 年ゼミ生 8 人
- 期間：11/4 から 11/25

2.2.2 課題内容

各週 2 問ずつ Python に関する内容の課題を与え、1 週間以内の提出期限を与え実施した。

課題内容は以下の通りで、検索すれば容易にソースコードを発見できるものと自ら考える必要のある問題を混ぜており、問題は回を追うごとに作成難易度を上げている。

- 11/4 から 11/10
 - 一般性を失わない塩分濃度に関する問題
 - 編集距離に関する問題
- 11/11 から 11/17
 - 三角形及び三角形の重心を描画する問題
 - 複素数に関する問題

- 11/18 から 11/25

- 線形代数に関する問題
- Python の文法に関する問題

また、問題を解くにあたり、ゼミ生内での協力のみとし、コピーアンドペーストを容認したうえで自由に解いてもらうことを条件とした。

2.2.3 データの取得方法

Git のローカルリポジトリにコミットするまでのコマンド、git log コマンドの結果を記録するコマンド、実行回数や実行時のファイルの内容等を記録する処理、及び Python 実行コマンドを抱合したスクリプトを、ファイル実行時に Python コマンドの代わりに実行してもらいデータの取得を行った。使用したログファイルは 3 種類あり、それぞれ、実行回数の記録、ファイル内容の記録、git log コマンドの結果を出力している。

実験終了後、対象者に 3 種類のログファイルを課題と共に提出してもらった。

2.2.4 評価

有効なデータ数は 8 人中 5 人であった。また、得られたログデータを元に csv ファイルを作成し、データ分析を行った。

学生ごとの各ファイルの実行回数、問題ごとの各学生のプログラムの類似度の遷移、Git から得られた追加数と削除数の遷移等をグラフ化したところ、不適切な学習行動を取っていたと思われる学生の発見に成功した。

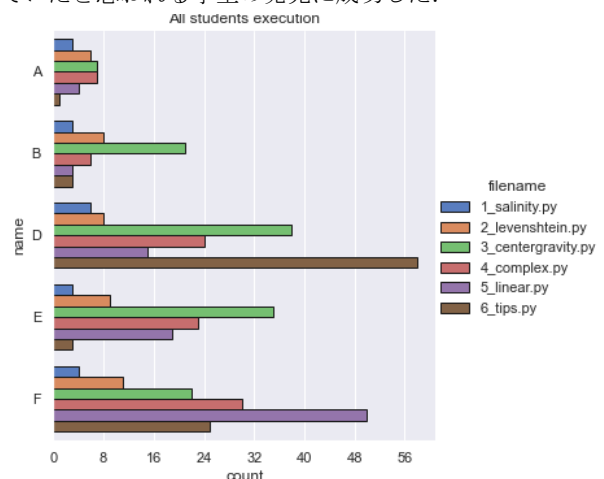


図1 学生ごとの各ファイルの実行回数

図1, より、問題難易度の上昇に対して明らかに実行回数の少ない学生がいることが分かるが、これだけでは不適切

な学習行動かどうかの判断はできない。

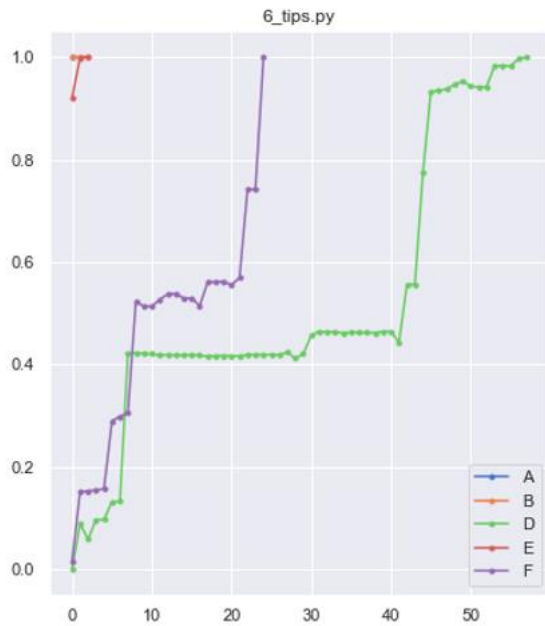


図2 各学生のプログラムの類似度の遷移 (問題6)

図2は提出したファイルの内容と実行ごとのファイルの内容を、レーベンシュタイン距離を用いて導出しており、最後が1.0になるように描画している。横軸は実行回数を表示している。

ここから、実行回数が少ないにもかかわらず類似度がすぐに高くなる学生や、最初から類似度が高い学生も存在していることが分かる。

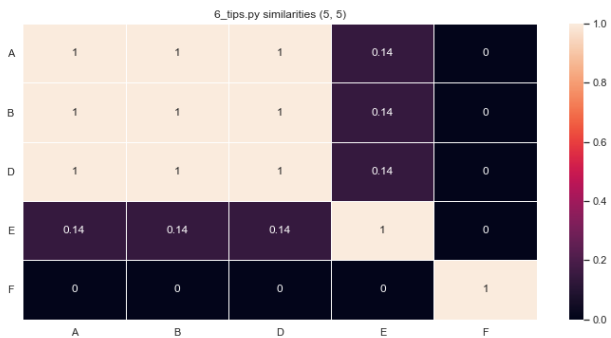


図3 抽象構文木の構造に着目した学生間の類似度 (問題6)

図3は提出したプログラムを抽象構文木に変換し、それぞれの構造に着目するため、集合の類似度で使用するDice係数とSimpson係数を用いて類似度を算出している。対角成分は自分同士の類似度のため1となっているが、対角成分以外でも1があることから、誰のプログラムをコピーアンドペーストしたのかが、図2と合わせて判明する。

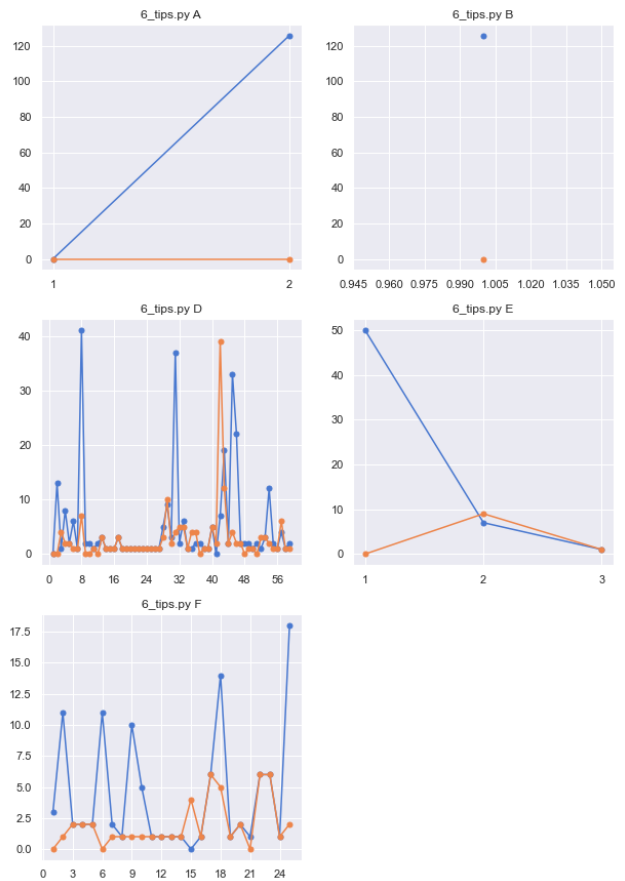


図4 各学生の追加数と削除数の変化 (問題6)

横軸は実行回数、縦軸は追加数及び削除数を表示している。図4のGitから得られた追加数と削除数の変化から、明らかに少ない実行回数に対して他の学生と比較して130近くの異常な数値が検出されている。

したがって、図1から図4までのグラフから不適切な行動を取っていると思われる学生の発見ができる。

3. おわりに

本研究では、不適切な学習行動を取る学生の発見が可能であり、その手法も従来のツールで十分可能であることも判明した。しかし、難易度の高い課題であれば今回のように不適切な学習行動を取る学生を発見することは容易だが、難易度が易しいものならその発見は上記の方法を以ってしても困難であることと考えられる。また、プログラムの類似度にレーベンシュタイン距離を利用したが、プログラムを文字列と考えた場合の類似度のため、正確なプログラム同士の類似度とは言えず、抽象構文木に着目した類似度に関しても変数名やprint文の中身といった細かい部分を考慮していないため、より厳密な類似度の算出するには様々な要素を考慮していく必要がある。

今後は本研究を一つのシステムにし、上記に挙げた問題点を回復させる必要がある。また、今回行わなかった人間関係に着目した分析を加えることで不適切な学習行動をとる学生の発見に説得力を持たすことができると考えている。