

Scratch プログラムの可視化による類似度推定

2016 年度 来住ゼミ
G13908 岩科智彩 G13924 森下汐美

2017 年 1 月 31 日

1 はじめに

近年プログラミング教育が推進され、日本でも小学校での導入が検討されている。子供のプログラミング教育の方法として近年よく利用されているのが米国マサチューセッツ工科大学のメディアラボが開発した Scratch である。Scratch とは GUI 環境を提供し、ブロックを組み合わせることでプログラムを作る ([1])。初心者にとっては使いやすい設計となっているため米国では利用者が増えている。また利用者同士ではプログラムを公開し共有するためのサイトも提供している。そこで実際にサイトで公開されている Scratch プログラムを利用して、プログラミング教育を支援するツールの開発を目指すことにした。

Scratch ではあるプログラムが引用された場合、その関係をリミックスツリーという図で表している。これでは引用関係は分かるが実際に引用されたもの同士の類似度は表されていないため、一箇所の変更を加えたもの、多数の変更を加えたものはリミックスツリー上では引用元のプロジェクトから同距離で表されている。この距離が数値化され、分かりやすく可視化できれば、例えば生徒の課題が提出された際にサンプルプログラムとの類似性で評価が分かりやすくなるのではないかと考える ([2])。本研究では Scratch プログラムを解析して類似度の数値化を行い、より分かりやすく可視化することを目指す。

2 研究するにあたって

本研究では収集データを元に類似度を数値化し、可視化を行ったのち類似性を評価する。この”類似度”と”類似性”の定義として、先行研究である「プログラム間の類似性の定量化手法」 ([3]) を参考に、”類似度”は実際にデータを計算式に当てはめ、数値として表示させたものを指し、”類似性”は最終的に可視化した結果が正しいのか、主観的に判断したものを指すことにする。

3 提案する手法

本研究では元プロジェクトと各プロジェクトをすべてのブロックが記録された 1 つの辞書型データを使用し、類似度推定を行う。類似度尺度は cos 類似度を用いる ([5])。cos 類似度とはベクトル空間モデルにおいて、文書同士を比較する際に用いられる類似度計算方法である。

収集できる Scratch プロジェクトのデータは json 形式 ([4]) であるため、Python 環境のもと json モジュールをインポートすることでファイルを読み込む。Scratch では数多くのブロックが用意されており、どのブロックを使用するかによって、まったく異なるプログラムを作成することができるため、本研究ではブロックの種類ごとの個数を集計をした後、ソートしたものをベクターに直すことで類似度計算を行う。また、使用されているブロック数とスプライト (オブジェクト) 数も類似度を推定する際の要素として、使用する。

【cos 類似度】

$$\cos(A, B) = \frac{\vec{A} * \vec{B}}{|\vec{A}| |\vec{B}|} = \frac{\vec{A} \cdot \vec{B}}{|\vec{A}| |\vec{B}|} = \frac{\sqrt{\sum_{i=1}^{|V|} A_i^2 * \sum_{i=1}^{|V|} B_i^2}}{\sqrt{\sum_{i=1}^{|V|} A_i^2 * \sum_{i=1}^{|V|} B_i^2}} \quad (1)$$

4 実験方法

本研究では 2 万弱の引用がされている”Pong Starter” (卓球ゲーム) の大規模なりミックスツリー ([6]) と 200 弱の引用がされている”Maze Game” (迷路ゲーム) の小規模なりミックスツリー ([7]) を利用し、元プロジェクトと各プロジェクトを比較をした。使用されているブロック・スプライトを数え上げ、類似度を表す cos 類似度を数値化し、これらのデータを使用してグラフ化する。元のプログラムから引用されたものを 1 段目、以後 1 段目のプロジェクトから引用されたものを 2 段目、2 段目から引用されたものを 3 段目とし、各段ごとにグラフを作成する。

5 実験結果

出力されたデータを組み合わせてさまざまな可視化を行った。散布図では縦軸に \cos 類似度の値、横軸に出力されたブロック数（スプライト数）を元のプログラムの数値で割った値をとり、“Pong Starter”の場合はさらに対数値をとる。グラフ上では（1,1）が最も類似しているプロジェクトであるため、この点から同距離のものを似ているものから順に青、赤、緑、紫、水色と区別してプロットをする。カラーマップでは各要素の個数を集計し、頻度の高いものから色別でプロットした。

5.1 Pong Starter

5.1.1 ブロック数と \cos 類似度の散布図

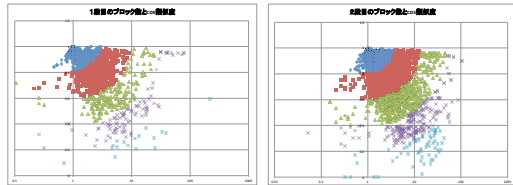


図 1 1 段目のグラフ

図 2 2 段目のグラフ

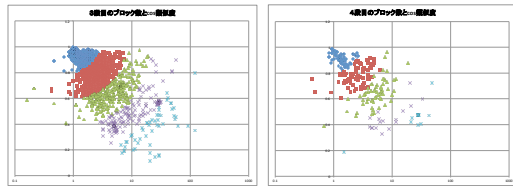


図 3 3 段目のグラフ

図 4 4 段目のグラフ

5.1.2 スプライト数と \cos 類似度の散布図

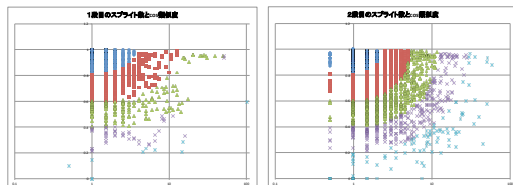


図 5 1 段目のグラフ

図 6 2 段目のグラフ

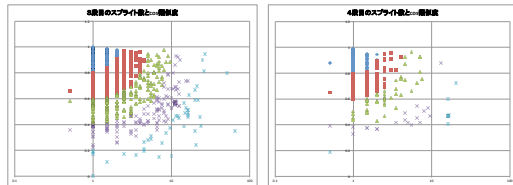


図 7 3 段目のグラフ

図 8 4 段目のグラフ

5.1.3 ブロック数と \cos 類似度のカラーマップ

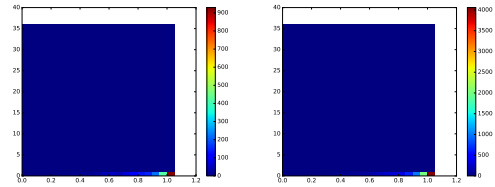


図 9 1 段目のグラフ

図 10 2 段目のグラフ

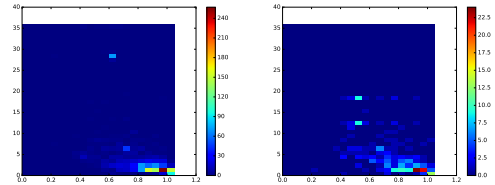


図 11 3 段目のグラフ

図 12 4 段目のグラフ

5.1.4 スプライト数と \cos 類似度のカラーマップ

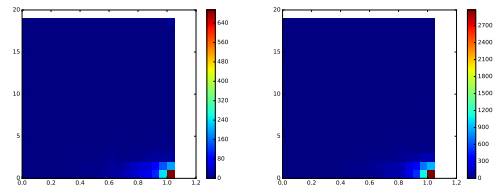


図 13 1 段目のグラフ

図 14 2 段目のグラフ

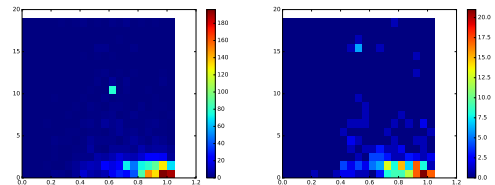


図 15 3 段目のグラフ

図 16 4 段目のグラフ

5.2 Maze Game

5.2.1 ブロック数と cos 類似度の散布図

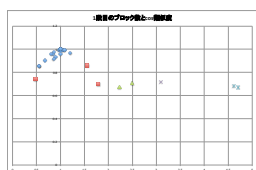


図 17 1 段目のグラフ

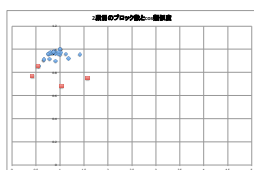


図 18 2 段目のグラフ

5.2.2 スプライト数と cos 類似度の散布図

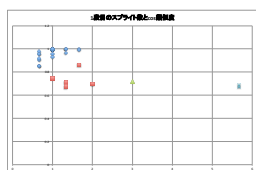


図 19 1 段目のグラフ

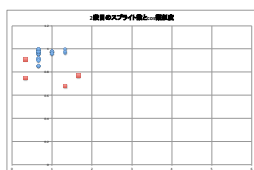


図 20 2 段目のグラフ

6 評価

6.1 方法 1

作成したグラフ上の点をいくつかピックアップし、その点にあるプロジェクトを実際に動かして元プロジェクトとの類似性を操作・音・外装・プログラムの動き・ブロックの木構造・ゲーム終了時の 6 つの項目で比較してみたところ、それぞれの段で最も似ているもの・最も離れているもの・中間という条件で取り出したプロジェクトは、やはりよく似ている・違う・少し違う内容となっていた。よって、今回使用したブロック数・スプライト数・cos 類似度はプロジェクト同士の類似度を図る上で重要な要素であったと言える。

また、1 段目から 4 段目まですべてを通して見てみると、2 段目以降にも元プロジェクトと同じような内容のものが含まれていることが分かった。これは、リミックスツリーの根元（元プロジェクト）から離れたプロジェクトも、元プロジェクトと同じ場合が生じているということが言え、リミックスツリーではその事実までは可視化されていなかったということが伺える。本研究で行った可視化手法では、例えば元プロジェクトへいくつも遡るところに位置するプロジェクトであったとしても、ほぼ正確な類似関係が可視化できており、新たな可視化として提案できるのではないかと考えている。ここに本研究の新奇性を見出すことができた。

6.2 方法 2

アンケート形式で類似性の相関を調査した。元プロジェクトと各プロジェクトを実際に動かし、5 段階評価をしてもらった。方法 1 と同様、可視化して見えてきた

類似度の位置関係と主観的な類似性はほぼ同じであることが分かった。

7 まとめと今後の課題

この可視化がツール上に加わることで、それぞれの生徒がどの程度自分のアイデアでプロジェクトを作成できているか、もしくはできていないのかという新たな情報を提供でき、教育者にとって個々の生徒の進捗を把握する一つの手助けになるのではないかと考えている。

本研究では 2 次元での可視化を行ったが、3 次元で可視化を行うことで、ブロック数・スプライト数・cos 類似度の関係性がより見え、誤差も少なくなると考えられる。また、ブロックの木構造を数値化できれば、類似度の精度が上がるのではないかと考えている。

参考文献

- [1] Scratch - Imagine, Program, Share, MIT Media Lab <https://scratch.mit.edu/>
- [2] 森秀樹・杉澤学・張海・前迫孝憲 (2011) 「Scratch を用いた小学校プログラミング授業の実践～小学生を対象としたプログラミング教育の再考～」
- [3] 小田悠介、若林茂 (2013) 「プログラム間の類似性の定量化手法」, [online] <http://www.kobe-kosen.ac.jp/activity/publication/kiyou/Kiyou12/Data/Vol51Paper103108.pdf>
- [4] 【Python 入門】JSON 形式データの扱い方 (2016 年 12 月 12 日), Morio <http://qiita.com/Morio/items/7538a939cc441367070d>
- [5] コサイン類似度 <http://www.cse.kyoto-su.ac.jp/g0846020/keywords/cosinSimilarity.html>
- [6] Remix tree for Pong Starter <https://scratch.mit.edu/projects/10000036/remixtree/>
- [7] Remix tree for Maze Game <https://scratch.mit.edu/projects/2768574/remixtree/>