

学習履歴に基づいたプログラミング学習者の動機付けの評価

富田 直輝[†]大枝 真一[‡]木更津工業高等専門学校 制御・情報システム工学専攻[†]木更津工業高等専門学校 情報工学科[‡]

1. はじめに

2020 以降日本国内において小学校から高等学校を対象にプログラミング教育が必須になる。プログラミング教育の狙いとしては IoT やビックデータ, AI などの第四次産業革命に対して論理的思考を備え対応できる人材の育成である。

また, 先行研究において, 木更津市内の小中学校の児童・生徒を対象に行った英語学習に関するアンケートの結果について次元削減を用いて可視化を行い, 動機付けの可視化を行った。このアンケートは 3 年間分行われており学年の移りがわりごとに動機付けが変わっていることを次元削減の手法である主成分分析や t-SNE[3] を用いることによって確認することができた。しかし, 個人を特定するものがなかったため追従して学習者の動機付けの移り変わりを見ることができなかった [1]。

そこで上記の 2 点を踏まえて, 本研究においては木更津高専情報工学科の学生を対象にプログラミングに関するアンケートを行い動機付けの可視化やグルーピングを行うことを提案する。木更津高専においては 15 歳からプログラミング教育を 5 年間行なっているため, 前述したプログラミング教育のモデルになりうると考えている。プログラミング学習者の動機付けを把握することは学習を効率に行うことに繋がると考えている。また, 自身が所属する高専で行うことによって個人の紐付けを行え, 1 年間において学習者の動機付けがどのように変わったかを調査できる。

プログラミングを対象者に行なったアンケートを次元削減を用いて可視化を行うことによって学生ごとにクラスを抽出し, 上記の課題を解決する。

2. UMAP

UMAP(Uniform Manifold Approximation and Projection) とは次元削減の手法の一つである。UMAP はデータを多様体空間で仮定することで次元削減を行う。ここで各データ点同士の関係はファジーの概念であるメン

バシッパ関数を与えることで表現することができる。

$$\mu : Any \rightarrow [0, 1] \quad (1)$$

式 (1) はあるメンバシッパ関数において μ が 0 に近くなるとデータが集合に属さず, 1 に近くなるとデータが集合に属していることを意味する。

次元削減前のデータ間は局地的な fuzzy simplicial 集合のメンバシッパ関数を $d(x, y)$ によって表現することで式 (2) で表される。ここで σ_i は最近傍の距離とし, ρ_i は正規化係数とする。

$$p_{j|i} = \exp[d(x, y) - \rho_i] / \sigma_i \quad (2)$$

また, p_{ij} は S ノルムを用いた論理和を用いて式 (3) 表される。

$$p_{ij} = (p_{i|j} + p_{j|i}) - p_{j|i}p_{i|j} \quad (3)$$

次元削減後の類似度は式 (4) で示される。ここで a と b は任意の係数として自由に決めることができる。

$$q_{ij} = (1 + a\|y_i - y_j\|_2^{2b})^{-1} \quad (4)$$

コスト関数は式 (3) と式 (4) を用いて式 (5) で表される。

$$C(p_{ij}, q_{ij}) \triangleq \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} + \left(1 - q_{ij} \log \frac{1 - p_{ij}}{1 - q_{ij}} \right) \quad (5)$$

$C(p_{ij}, q_{ij})$ は t-SNE と同じく最急降下法を用いることによって更新式が求められる。

UMAP は t-SNE と違い, 次元削減前の類似度を自由に指定することができるのでデータによって柔軟に変えることができることが利点となっている。また, 次元削減後の初期値は k 近傍探索によって得られた重み付き巡回行列を与えることになる [2]。

3. 計算機実験

本校の情報工学科 2 年生を対象にプログラミングの授業に関するアンケートについて 6 月から 12 月に渡ってアンケートを 5 回行った。アンケートの内容は指導教員と相談を行いつつ, その日の授業に関連づけたアンケートを行った。抜粋したアンケートの内容は表 1 に示される。なお, 便宜上 1 回目から 5 回目は A~E で表すことにする。

Rating motivation of programing beginner based on learning log.

[†]Naoki TOMITA · Advanced Course of Control and Information Engineering, National Institute of Technology, Kisarazu College

[‡]Shinichi OEDA · Department of Information and Computer Engineering, National Institute of Technology, Kisarazu College

表 1: 質問の内容 (一部抜粋).

項目番号	質問の内容
A0	プログラミングを学ぶのは楽しい.
A4	将来研究室に入った時に必要になるので勉強している.
B1	ペアプログラミングはスキル向上に役立つ.
B8	自宅にプログラミングを行う環境がある.
B9	家でプログラミング学習を行う時間(平均)を教えてください.
C1	期末試験の勉強はもう始めている.
C11	授業が楽しみだ
D0	プログラムの仕様をノートなどに書き出してからコーディングしている
D6	自宅でのプログラミング環境を構築したことがある
E0	自分で書いたプログラムが予想通りに動いたときに達成感を感じる.
E3	ゲーム制作のプログラミングは楽しい.
E5	プログラミングは楽しい.

表 2: 各クラスターにおける上位 5 項目.

0(星)	1(四角)	2(円)	3(三角)
B6	B8	A7	E0
B8	E0	B8	A7
A0	D6	D6	B0
A4	D0	B1	B1
B5	E5	E0	B9

また上記のアンケートによって得られた結果を元に UMAP を用いて可視化を行い, DBSCAN[4] を用いることによってクラスタリングを行った.

4. 結果・考察

上記のアンケート結果によって得られた結果に対して, 回答者が少なかった第 3 回目の結果と 1 回以上アンケートの回答を行っていない学生が含まれるデータを除いて可視化を行った. これによって対象となるデータ数は 35 個である.

UMAP を行い DBSCAN を用いてクラスタリングした結果を図 1 に示す. また, UMAP におけるクラスターごとの類似性が高い項目を分散が小さい順にソートしたものを表 2 に挙げる. なお, クラスターの振り分けは図 1 の各点ごとの形に従って示される.

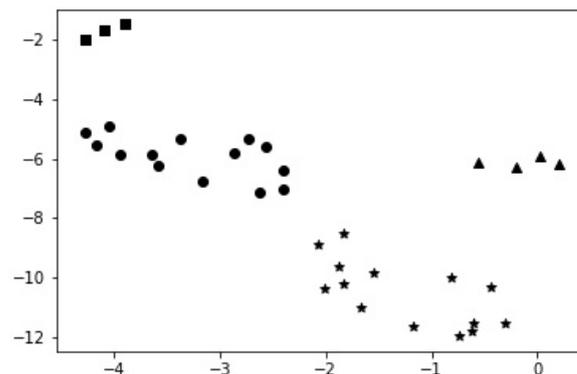


図 1: UMAP による可視化の結果.

表 2 においてクラスター 0, クラスター 2, クラスター 3 は動機付けが高い学生であることが考えられる. クラスター 1 においては E0 と E5 の回答の平均が 5 であり, そう思わないに該当する回答であることから動機付けが低い集団であることが考えられる.

5. まとめ

本研究においては次元削減の手法である UMAP を用いることによって可視化を行うことができた. また, クラスターを抽出することで動機付けが低い集団の特定を行うことができた. 今後の課題として, アンケートから得られた結果以外にもキーログなどの学習履歴を特徴量とし, 可視化を行うことが挙げられる.

謝辞

本研究は, JSPS 科研費 19H01728 の助成を受けたものです.

参考文献

- [1] 富田直輝, 大枝真一, Benjamin Maynard, “次元削減を用いた小中学校英語学習者間の動機付けの比較”, 2019 年電子情報通信学会総合大会, D-15-40, 2019
- [2] Leland McInnes, John Healy, James Melville, “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction”, arXiv:1802.03426, 2018
- [3] Laurens van der Maaten, Geoffrey Hinton, “Visualizing Data using t-SNE”, Journal of Machine Learning Research 9, pp.2579-2605, 2008
- [4] Martin Ester, Hans-Peter Kriegel, Jiirg Sander, Xiaowei Xu, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”, Proceedings of the second ACM International Conference on Knowledge Discovery and Data Mining (KDD), 226-231, 1996