

次元圧縮を用いた小中学校英語学習の内発的・外発的動機の評価

Evaluation Intrinsic and Extrinsic Motivation of English Learners in Elementary and Junior High School by Dimension Reduction Method

富田 直輝[†]
Naoki Tomita

大枝 真一[‡]
Shinichi Oeda

Benjamin Maynard[§]
Benjamin Maynard

1. まえがき

近年、国際社会において英語の習得は重要になってきており、特に小学校や中学校からの早期英語教育が必要となっている。現時点では5年生6年生においては外国語活動で英語学習が行われているが、2020年には小学校の3,4年生の外国語活動,5,6年生においては英語が教科化されることになっている。そのため、英語学習の環境の質を高めるためにも評価が重要になってくると考えられる。

我々の先行研究において、英語学習の動機は「海外への興味」、「内発的動機付け」、「家族および近親者の励まし」、「外発的動機付け」、「不安」の5つの要素に分けられると仮定した [1]。その上で小学校5年生から中学校3年生を対象に4件法のアンケートを行った。その結果、中学校1年生,2年生は英語学習に取り組むモチベーションが一番低くなった [1]。

しかしながら、先行研究ではデータ毎の平均を取って評価を行っているため個々のアンケート結果の可視化を行えていなかった。そこで本研究では主成分分析,カーネル主成分分析,t-SNEを用いて次元圧縮を行い可視化を行い学年毎の傾向について評価を行う。

2. 先行研究

英語学習の動機付けの先行研究はすでに行われている。例として小学校教育における国際交流を行うことによって、児童の内発的動機,外発的動機が交流前に比べて交流後では高まるといったことがアンケートの解析からわかったとしている先行研究がされている [2]。また、成人以上の日本人英語学習者を対象にした英語学習者の動機付けにおける専攻研究は「海外の旅行のため」、「視野を広げるためといった」といった結果を得ることができたとしている [3]。

英語学習以外の外国語学習における動機付けの研究においても外国語学習者の動機づけが研究されている。中国人留学生における日本語学習者においては学習者が目標言語の文化や興味があるといった統合的動機や将来の就職のためであるといった道具的動機の組み合わせであるとしており、日本人大学生における中国語学習者においては他からの影響が要因とする誘発的動機,統合的動機であるとした [4]。

アンケートの解析の観点において解析対象となるア

ンケートの形式は自由記述式回答のアンケートと選択式回答アンケートの2つが挙げられる。自由記述式アンケートについてはテキストマイニングによる解析が行われている。また、選択式アンケートにおいては主成分分析,因子分析,多次元構成尺度法を用いた解析が主な手法となっている。また、他の手法としては自己組織化マップを用いたクラスタリング,およびアンケートの可視化が行われている。

教育分野におけるアンケートの結果を用いた研究においては多次元構成尺度法や主成分分析を用いた可視化及び分析が行われている [5,6]。教学支援システムについては選択式回答と自由記述式回答の2つの回答方法をとっている。選択式アンケートについては主成分分析を用いて特徴抽出を行っており自由式アンケートについてはクラスタリング分析を用いている [5]。学生に対するアンケートの解析においては因子分析を用いて解析を行い,ISOMAPを用いて学生の成績と結び付けクラスタリング及び可視化を行っている [6]。

教育分野以外における可視化の例としては福岡県の水質調査のアンケートの結果を可視化するために自己組織化マップを用いて年齢や住居別の観点から考察を行っていた [7]。

これまでに紹介した英語学習の動機付けの研究とアンケートの可視化の研究はそれぞれ行われているが、我々は小中学校の生徒を対象とし、可視化技術としてt-SNEを用いて高次元データの可視化を試みる。

3. 主成分分析

主成分分析(以下,PCA)とは統計データから互いに無関係の因子を取り出して、観測対象をそれらの因子線型結合で説明することであり、取り出された因子を主成分と呼ぶ。それらの分散(対角に変換された分散共分散行列の対角要素)は各々の主成分の寄与の程度を表している [5]。

ここで p 個の次元で表すことができる n 個 ($i = 1, 2, \dots, n$) のデータに基づいて分散共分散行列を表すと式 (1) で表される。 \bar{x} はデータ数 n の平均である。

$$S = \begin{pmatrix} s_{11} & \cdots & s_{1k} \\ \vdots & & \vdots \\ s_{j1} & \cdots & s_{jk} \end{pmatrix} \quad (1)$$

$$s_{jk} = \sum_{i=1}^n (x_{ij} - \bar{x}_i)(x_{ik} - \bar{x}_k) \quad (2)$$

$(j = k = 1, 2, \dots, n)$

この分散共分散行列 S について固有値と固有ベクトルを $\lambda - pI = 0$ を求める。最大固有値を λ_1 として固有

[†]木更津工業高等専門学校 制御・情報システム工学専攻, Advanced Course of Control and Information Engineering, National Institute of Technology, Kisarazu College

[‡]木更津工業高等専門学校 情報工学科, Information and Computer Engineering, National Institute of Technology, Kisarazu College

[§]神田女学園中学校高等学校, Kanda Girls' Junior-Senior High School

値を

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0 \quad (3)$$

とする。また単位ベクトルは

$$w_1 = \begin{pmatrix} w_{11} \\ w_{12} \\ \vdots \\ w_{1p} \end{pmatrix} \quad w_2 = \begin{pmatrix} w_{21} \\ w_{22} \\ \vdots \\ w_{2p} \end{pmatrix} \quad w_p = \begin{pmatrix} w_{p1} \\ w_{p2} \\ \vdots \\ w_{pp} \end{pmatrix} \quad (4)$$

となる。固有値と固有ベクトルの行列は次の式で与えられる。

$$A = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \end{pmatrix} \quad (5)$$

$$W = (w_1, w_2, \dots, w_p) \quad (6)$$

また p 次元のデータは次の式で変換することができる。

$$\begin{aligned} y_1 &= w_{11}(x_1 - \bar{x}) + w_{12}(x_2 - \bar{x}) + \dots + w_{1p}(x_p - \bar{x}) \\ y_2 &= w_{21}(x_1 - \bar{x}) + w_{22}(x_2 - \bar{x}) + \dots + w_{2p}(x_p - \bar{x}) \\ &\vdots \\ y_p &= w_{p1}(x_1 - \bar{x}) + w_{p2}(x_2 - \bar{x}) + \dots + w_{pp}(x_p - \bar{x}) \end{aligned} \quad (7)$$

2次元間にプロットする場合は第2主成分までが軸でプロットされることになる。このことを次元圧縮という。 i 成分における情報量を知るための寄与率 ϕ は次の式で与えられる。

$$\phi = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad (8)$$

また第1主成分から第 k 主成分までを用いたときの情報量を知るための累積寄与率 Φ は次の式で与えられる。

$$\Phi = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad (9)$$

4. カーネル主成分分析

通常の PCA はデータを直接線形変換して分散が大きくなるような射影を求めるため、直線や平面といった線形の部分構造しか取り出すことができない。その点カーネル主成分分析 (以下、KernelPCA) は高次元のベクトルに特徴変換を行ってから主成分分析を行うことによって低次元の線形部分空間を求めるため、特徴を保ったまま変換を行うことができる。

平均が 0 になる場合、分散の式は次の式で表される。

$$\text{Var}_n[f(\mathbf{x})] = \frac{1}{n} \sum_{i=1}^n (\mathbf{w}^\top \phi(\mathbf{x}^{(i)}))^2 \quad (10)$$

ラグランジェの未定乗数法を用いて最適化を行うことにすると式は次で表される。

$$L(\mathbf{w}) = -\text{Var}_n[f(\mathbf{x})] + \lambda(\|\mathbf{w}\|^2 - 1) \quad (11)$$

最終的な式は次で表される。

$$K\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha} \quad (12)$$

なお $\boldsymbol{\alpha}$ は α_i を次のように定めたベクトルである。

$$\alpha_i = \frac{\mathbf{w}^\top \phi(\mathbf{x}^{(i)})}{n\lambda} \quad (13)$$

つまりグラム行列の固有問題を求めることによって λ が求まるとしている。固有ベクトルは $\boldsymbol{\alpha}^\top K\boldsymbol{\alpha} = 1$ の制約を満たすように $\boldsymbol{\alpha}$ を求めれば良い。また、2次元以上の空間に射影する場合は、上から M 個の固有値を取り、対応する固有ベクトルによって張られる空間に射影を取れば良い。

平均が 0 でない場合、「分散 = 二乗平均 - 平均の 2 乗」の関係からサンプルの分散を導出する。サンプル平均は次の式で表される。

$$\begin{aligned} E_n[f(\mathbf{x})] &= \frac{1}{n} a_i k(\mathbf{x}^{(i)}, \mathbf{x}^{(l)}) \\ &= \frac{1}{n} \boldsymbol{\alpha}^\top K\mathbf{1} \end{aligned} \quad (14)$$

$\mathbf{1}$ は 1 を n 個並べた $(1, 1, \dots, 1)^\top$ というベクトルである。平均の 2 乗は次の式で表される。

$$E_n[f(\mathbf{x})] = \frac{1}{n^2} (\boldsymbol{\alpha}^\top K\mathbf{1})(\mathbf{1}K\boldsymbol{\alpha}^\top) \quad (15)$$

二乗平均から式 (15) を引くと次の式で表される。

$$\text{Var}_n[f(\mathbf{x})] = \frac{1}{n} \boldsymbol{\alpha}^\top K J_n K \boldsymbol{\alpha} \quad (16)$$

ただし J_n は次の行列で表される。

$$J_n = I_n - \frac{1}{n} \mathbf{1}\mathbf{1}^\top \quad (17)$$

したがって次の固有値問題を解く式で求めることができる。

$$J_n K \boldsymbol{\alpha} = \lambda \boldsymbol{\alpha} \quad (18)$$

5. t-SNE

t-SNE(t-distributed Stochastic Neighbor Embedding) とは高次元上におけるユークリッド距離をデータ点の類似度を表現する条件付き確率に変換する手法になっている。高次元空間についてはガウシアン分布を用い、低次元空間においては自由度 1 の t-分布を用いる。条件付き確率 $p_{j|i}$ と同時確率 q_{ij} は式 (19)、式 (20) で表される。

$$p_{j|i} = \frac{\exp(-|x_i - x_j|^2 / 2\sigma_i^2)}{\sum_{k \neq x} \exp(-|x_i - x_k|^2 / 2\sigma_i^2)} \quad (19)$$

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N} \quad (20)$$

自由度 1 の t-分布を用いた同時確率 q_{ij} は式 (21) で表される.

$$q_{ji} = \frac{1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}} \quad (21)$$

t-SNE はユーザーが定めた perplexity によって σ_i の値について二分探索を行う. perplexity の式 (22) 与えられる.

$$\text{Perp}(P_i) = 2^{H(P_i)} \quad (22)$$

式 (22) における $H(P_i)$ は式 (23) で与えられる.

$$H(P_i) = - \sum_j p_{j|i} \log_2 p_{j|i} \quad (23)$$

perplexity は x_i における近傍に対する効果的な数を解釈することができる. perplexity は一般的に 5 から 50 に設定する.

コスト関数を最小にするような方程式是最急降下法を用いることとし, 式によって得られる勾配は次のように与えられる. データ点 \mathbf{P} とマップ点 \mathbf{Q} の KL 距離の勾配の式は式 (24) で表される.

$$\frac{dC}{dy_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)(1 + \|y_i - y_j\|^2)^{-1} \quad (24)$$

$(y_i - y_j)$ はマップ点における距離を表現し, $(p_{j|i} - q_{j|i} - p_{i||j} - p_{i||j})$ はマップ点における類似度とデータ点における類似度の不整合性を表すものとなっている.

更新式は最急降下法で求められ, 式 (25) で表される.

$$\mathbf{Y}^{(t)} = \mathbf{Y}^{(t-1)} + \eta \frac{dC}{d\mathbf{Y}} + \alpha(t)[\mathbf{Y}^{(t-1)} - \mathbf{Y}^{(t-2)}] \quad (25)$$

式 (25) において $\mathbf{Y}^{(t)}$ は t におけるマップ点を表し, η は学習率, $\alpha(t)$ は t における慣性を表す.

6. 計算機実験

2015 年に木更津市内の小中学校の児童生徒 702 人を対象に行ったアンケート結果を用いて PCA, KernelPCA, t-SNE を行う. アンケートの項目は 19 個あり, 内容は表 1 になる. アンケートの回答の答えは 1 から 4 のうちの 1 つを選び, 1 がそう思う, 4 が全くそう思わないとなっている. またアンケート項目は表 2 で分類される.

7. 結果・考察

対象のデータについて主成分分析, カーネル主成分分析, t-SNE を行った結果を図 1 から図 3 示す. 色分けについては小学校 5 年生が青色, 小学校 6 年生が赤色, 中学校 1 年生が緑色, 中学校 2 年生は黄色, 中学校 3 年生となっている.

図 1 の PCA の軸については第 1 主成分については負の値が大きくなるほど両方の動機付けが強くなって

表 1 英語学習に関するアンケートの項目.

1	英語学習は楽しい.
2	いろいろな外国に行ってみたい.
3	英語の授業がある日は楽しみである.
4	外国の友達をたくさん作りたい.
5	英語の授業では, みんなよりできないんじゃないかと心配になる.
6	習った英語をもっと使ってみたい.
7	中学校に入っても困らないように勉強をしているつもりだ.
8	英語の授業は, もっと多い方がいい.
9	うちの人は, 英語の勉強がとても大切だと思っている.
10	英語の授業中は, いつも, 何となく心配である.
11	大人になったら必要だと思うので, 英語を勉強している.
12	英語が上手になって, 外国の人と話してみたい.
13	うちのひとは, 私が英語ができるように望んでいる.
14	将来なりたい仕事のために英語を勉強している
15	外国に住んでみたい
16	英語の授業で, 答えたり, 表することはドキドキする
17	外国のことをもっと知りたい
18	うちの人から, 英語を一生懸命勉強しろと言われる.
19	高校や大学に入るために必要なので, 英語を勉強している.

表 2 アンケートの項目の分類.

分類	対応する設問
内発的動機	1, 2, 3, 4, 6, 8, 12, 15, 17
外発的動機	7, 9, 11, 13, 14, 18, 19

いた. 第 2 主成分については正の値が大きくなるほど内発的動機が強くなり, 負の値が大きくなるほど外発的動機が強いと考えられる.

図 2 の KernelPCA の結果においては軸は第 1 主成分においては負の値が大きくなるほど 1 を多く答えており, また正の値が大きくなるほど 3 を多く答えていた. また第 2 主成分においては負の値が大きくなるほど 2 を多く答えており, また正の値が大きくなるほど 4 を多く答えていた.

これは PCA を適用する前にカーネル関数を用いて高次元に変換を行っているため構造が変わってしまい本来の目的である内発的動機・外発的動機を検討する手法には向いていないと言える.

図 3 の t-SNE の結果の解析には k-means 法を用いて解析を行った. その結果については主成分分析と同じように X 軸においては内発的動機と外発的動機の両方を総合したような軸であり, Y 軸においては内発的動機と外発的動機のどちらに影響されているかを示すような軸であると考えられる. また t-SNE の性質の通り近いデータ同士が近いところに配置されていること

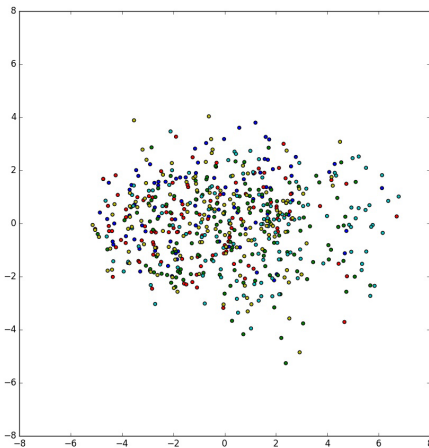


図1 PCAにおけるアンケートの分析結果.

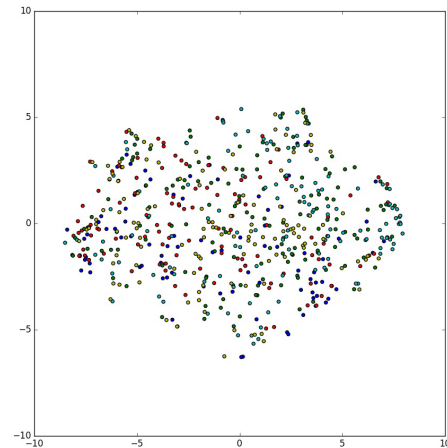


図3 t-SNEにおけるアンケートの分析結果.

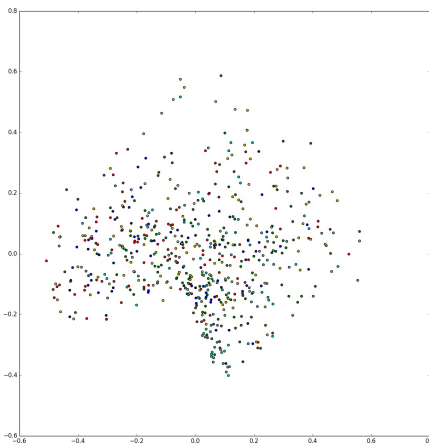


図2 Kernel PCAにおけるアンケートの分析結果.

が図3から考えられる。

以上から図1から図3よりPCAとt-SNEにおいては学年ごとの傾向を見ることができたと考えられる。学年ごとの比較においては中学校1年生、中学校2年生においては外発的動機が低くなっている。

8. まとめ

次元圧縮の手法であるPCA, t-SNEを用いることによって小学生中学生の英語学習者の内発的動機, 外発的動機に関して可視化を行うことができた。また, t-SNEの方がPCAより良いと考察することができた。しかしながら, 定量的な判断ができていないため課題であると考えられる。

今後の予定として, 2016年, 2017年分のアンケート結果を取得しているためそれらについてPCA, t-SNEを用いて2015年のアンケート結果と比較を行い検討を行う予定である。

参考文献

- [1] Benjamin Maynard, “Kisarazu City, Manabi Shien Center - Attitudes to ESL Learning Study”, kisarazu city, pp.1-13, 2016.
- [2] 山本淳子, “小学校英語教育における国際交流の役割と意義”, 新潟経営大学紀要 17 巻, pp.103-116, 2011.
- [3] 小磯かをる, “日本人英語学習者の動機付け—JGSS-2003のデータ分析を通して—”, 研究論文集 [4] JGSS で見た日本人の意識と行動, 2005.
- [4] 毛賀力, 福田 倫子, “中国における日本語専攻学習者及び日本における中国語専攻学習者の動機づけの比較”, 言語と文化 23 巻, pp.209-32, 2011.
- [5] 渡辺智幸, 後藤正幸, 石田嵩, 酒井哲也, 平澤茂一, “教学支援システムに関する学生アンケートの分析” 情報科学技術フォーラム一般講演論文集 4 巻 4 号, pp.317-318, 2005.
- [6] 志垣一郎, “学生による授業アンケートと成績の関係”, Memoirs of the Osaka Institute of Technology, Series A, Vol. 55, No.1, pp.1-9, 2010.
- [7] 横田いずみ, 井料隆太, 井芹慶彦, 広城吉成, 神野健二, “自己組織化マップを用いた福岡市民の水に関するアンケート調査結果分析”, 水工学論文集 53, pp.553-558, 2009.
- [8] Laurens van der Maaten, Geoffrey Hinton, “Visualizing Data using t-SNE”, JMLR Volume 9, pp.2579-2605, 2008.