

学習履歴データの属性を考慮した可視化による学習予測の検討

Guidelines for the Proceedings of the Annual Conference of JSiSE

小川 賀代^{*1}, ピトヨ ハルトノ^{*2}
Kayo OGAWA^{*1}, Pitoyo HARTONO^{*2}

^{*1}日本女子大学理学部

^{*1}Faculty of Sciences, Japan Women's University

^{*2}中京大学工学部

^{*2}School of Engineering, Chukyo University

Email: kayo@optnet.jwu.ac.jp

あらまし：本稿では、LMS から取得した学習履歴データを主成分分析、線形判別分析、SOM、CRSOM の4つの手法を用いて可視化を行い、結果を比較した。その結果、属性を考慮し、非線形写像によって可視化する CRSOM を用いると、他の手法に比べて、同じ属性のデータが集まり、異なる属性との分離する傾向が得られた。これにより、CRSOM を用いることで、学習到達度の予測に活用できる可能性が得られた。

キーワード：ログ解析、可視化、学習履歴データ、SOM

1. はじめに

学習に関するデータを大量に収集し分析して学習支援へ繋げる Learning Analytics が注目されている。その中で、e-learning の受講状況を示す学習履歴データを LMS (Learning Management System) から得て、受講者の学習到達度の予測に取り組んでいる。学習支援に活かすためには、予測結果だけでなく、その結果に至った根拠まで示せることが望ましい。しかし、SVM (Support Vector Machine)[1]等での識別問題としての予測では結論のみしか得られない。学習履歴データと到達度を示す属性を結びつけて可視化をすれば、同じ到達度で異なる学習傾向の人や、似た学習傾向で異なる到達度の人などの情報を直感的に把握可能と考えられる。また、個人の学習は多種多様のため、学習履歴データは非線形構造だと考えられる。そこで、本研究ではデータの属性も考慮して非線形関係を可視化する手法 CRSOM (Context-Relevant SOM)[2]を用いて、学習履歴データを可視化し、他の手法と比較検討した。

2. CRSOM によるデータの可視化

データの可視化へ向けた2次元縮約手法は様々あり、線形変換では、PCA (Principal Component Analysis)、属性を考慮する LDA (Linear Discriminant Analysis)、非線形変換では、SOM (Self-Organizing Map)[3]がある。CRSOM は、制限付き RBF ネットワーク (rRBF: restricted RBF Network) を用いて2次元に写像する方法である。rRBF は関数近似に用いられる RBF ネットワーク(Radial Basis Function Network)[4]の隠れ層ユニットに SOM のような近傍の概念を取り入れたものである。rRBF にデータを入力すると、2次元格子状に並んだ隠れ層のユニットの中からそれに最も近いものと周囲のユニットだけの情報を元に属性判別し出力する。rRBF の学習は、出力が入力データの

属性と一致するように各ユニットを修正することで行う。式は下記に示す。

$$W \leftarrow W + \eta \sigma \delta^h (X - W)$$

ここで、 η は学習率、 σ は近傍関数、 δ^h は修正方向を示す。図1に rRBF ネットワークの概念図を示す。入力 X と同じ次元のベクトル W を並べたマップを用意し、入力 X に最も近い参照ベクトルとその近傍を学習させる。学習終了後、各データをそれに近い隠れ層ユニットの位置にプロットしたものが CRSOM である。

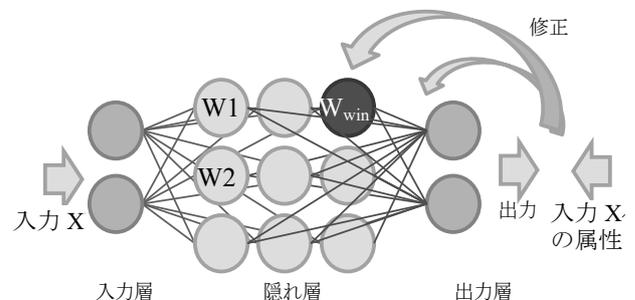


図1 rRBF ネットワークの概念図

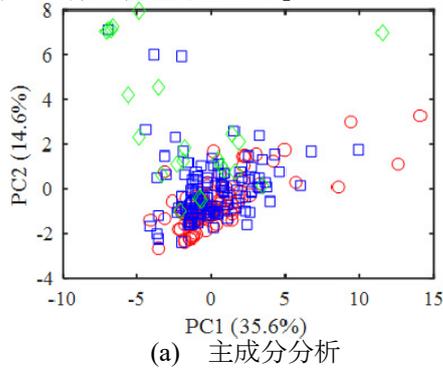
3. 学習履歴データへの適用

可視化の手法を比較するために、企業が提供する e-learning の LMS から 23 項目の学習履歴データ (10 章の各学習回数、章末テスト受験回数、章末テストの結果等) を 238 人分用いて解析を行った。属性は、学習到達度とし、修了テストの点数で A (90 点以上)、B (70~89 点)、C (69 点以下) の 3 つに属性付けした。

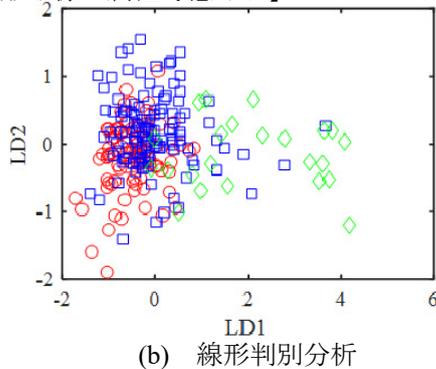
今回は、属性を考慮せず線形写像によって可視化を行う主成分分析 (PCA)、属性を考慮して線形写像によって可視化を行う線形判別分析、属性を考慮せず非線形写像によって可視化を行う SOM、属性を考慮して非線形写像によって可視化する CRSOM を用いて可視化し、比較を行った。その結果を図 2(a)~

(d)に示す. SOM は学習回数 2000, 学習率 0.01, CRSOM は学習回数 2000, 出力層の学習率 0.1, 隠れ層の学習率 0.1 とした. 図 2(c)(d)において, 一つのユニット上のデータが同じ属性のとき印を大きくし, 異なる属性のとき×をつけた.

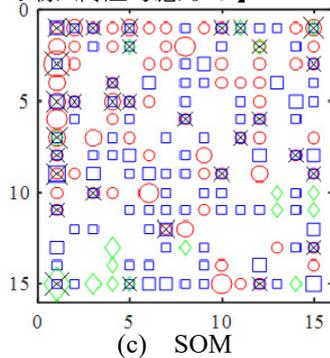
【線形写像・属性考慮なし】



【線形写像・属性考慮あり】



【非線形写像・属性考慮なし】



【非線形写像・属性考慮あり】

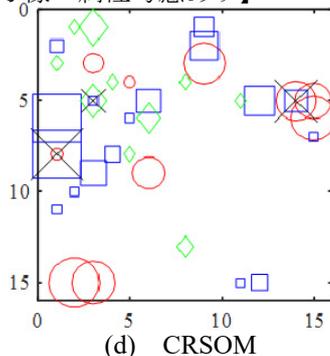


図 2 可視化手法の比較結果

図 2 より, 線形写像の(a)PCA, (b)LDA はデータが重なり属性の区別が難しいのに対し, (c)SOM, (d)CRSOM では非線形構造データに対応し分散したマップが得られた. (d) CRSOM は, 学習履歴データと到達度を結びつけて可視化することにより, マップ上で同じ属性のデータが近い位置に集まる結果が得られた. 更に, 他の LMS によって取得した学習履歴データに対しても適応したところ, 同様の結果が得られた.

4. まとめと今後の課題

学習履歴データを 4 つの手法 (属性考慮あり/なし, 線形写像/非線形写像) を用いて可視化を行い, 比較を行った. その結果, 属性を考慮し, 非線形写像によって可視化する CRSOM を用いると, 他の手法に比べて, 同じ属性のデータが集まり, 異なる属性との分離する傾向が得られた. これにより, 学習到達度の予測に活用できる可能性が得られた. 今後, 現在の可視化マップ上に, 新たな学習履歴データが入力された時の到達予測を実施し, 予測率の検討を行っていく.

参考文献

- (1) C. Cortes, V. Vapnik: "Support-vector networks", Machine Learning, 20, 3, 1573-0565 (1995)
- (2) P. Hartono, P. Hollensen, T. Trappenberg: "
- (3) Learning-Regulated Context Relevant Topographical Map", Neural Networks and Learning Systems, IEEE Transactions, 26, 10, 2323-2335, (2015)
- (4) T. Kohonen: " Self-organized formation of topologically correct feature maps", Biol Cybern, 43, 59-69 (1982)
- (5) T. Poggio, F. Girosi: "Network for Approximation and learning", Proceedings of IEEE, 78, 9, 1484-1487 (1990).

謝辞

本研究は, 文部科学省科学研究費補助金 (基盤研究 (C): 課題番号 25330419) の助成を受けたものである.