

平成 19 年 1 月 15 日

知識情報工学専攻	学籍番号	033730
申請者氏名	藤川 哲史	

指導教員氏名	村越 一支
--------	-------

## 論文要旨 (修士)

論文題目	Category Utility により自律的に成長する成長型階層自己組織化マップ
------	---

自己組織化マップ (Self-Organizing Map : SOM) は、多次元データを、データ間の位相関係を保持したまま、低次元のマップに写像することが可能な教師なしのクラスタリング手法の一つで、様々な分野で利用がなされている。しかし、SOM には、適切なマップサイズを事前に指定しなければ学習が正確に行われぬ問題と、入力データのクラスタが持つ階層構造を表現できない問題があった。

成長型階層自己組織化マップ (The Growing Hierarchical Self-Organizing Map : GHSOM) は、入力データに応じて、動的にマップサイズと、階層構造を拡張することで、SOM が持つ問題の解決を目指した改良型 SOM である。しかし、GHSOM にも二つの問題点がある。一つは、成長制御の基準尺度である量子化誤差は、マップの状態表現を行うには問題があり、マップが適切なサイズになるように、成長処理を制御出来ない可能性を持つことである。もう一つは、GHSOM における SOM を利用したデータの分類は、一部で過度な細分化が発生することがあり、その分類を基に階層化を進めていくと、階層構造が不必要に複雑化することである。一つめの問題は、量子化誤差がマップの状態、つまり、マップ上のデータの割り当てに対して、区別の明瞭さなどの基準で良さを評価ができないことに原因があり、量子化誤差に代わる指標が必要である。また、データの割り当て、すなわち量子化ユニットへのデータの分類に対して、正しく評価可能な指標があれば、二つめの問題で発生する、分類が過剰に細分化された部分を、その指標に従って補正できる可能性がある。

この、要求を満たす評価指標が、機械学習の一種である概念クラスタリング (Conceptual Clustering) において研究されている。概念クラスタリングでは、クラスタが持つ意味を評価する関数を用いて、意味解釈が容易なクラスタの構築と分類を行う。ただし、既存の概念クラスタリング手法は、SOM のように入力データの位相構造を学習することができない。

そこで本研究では、概念クラスタリングで利用されるクラスタ評価指標を、GHSOM の成長制御と、過剰な分類細分化を補正する処理に用いることで、GHSOM が持つ問題の解決した、位相構造を獲得可能なクラスタリング手法を提案する。評価指標には、Category Utility (CU) を用いた。CU は心理学研究をもとに、ヒトがもつ良い概念を予測するために開発された概念評価指標で、意味解釈が容易なクラスタ構築のための指標となることが期待できる。提案手法は、CU を基準として自律的に成長処理を制御し、細分化された分類を補正することで、クラスタの区分が明瞭な SOM マップの獲得、つまり次元圧縮・データ要約を行うとともに、階層的関係をもつ分類構造を適切に構築することが期待出来る。

提案手法と、従来の GHSOM との性能比較実験を、実データと、人工データを利用して行い、GHSOM によって入力データの位相構造が獲得可能なことを確認した。さらに、量子化誤差を用いた従来の GHSOM では、データによっては適切なマップサイズの獲得ができないこと、改良手法ではその問題が解決できており、かつ、クラスタ細分化が補正できたことを示した。