

1 T-O 4

自己組織化マップによる 臨床症例クラスタリングシステムの開発

—属性ベクトルの重み変化による分類効果—

納富 一宏¹⁾ 山口 俊光¹⁾ 斎藤 恵一²⁾ 藤本 哲男³⁾

¹⁾ 神奈川県立大学情報工学科 ²⁾ 東亜大学経営学部経営学科 ³⁾ 芝浦工業大学工学部機械工学科

1. はじめに

データベース化された臨床症例報告文書から、その疾患についての特徴的な語句や表現、さらには各種検査項目に関する情報を自然言語処理的に自動抽出して、症例の疾患系分類を目的とする自己組織化マップ (SOM: Self-Organizing Map) を用いた症例文書クラスタリングシステムの構築を行っている^{[1],[2],[3]}。

本稿では、自己組織化マップにおける属性ベクトルの重みを変化させた場合に、疾患系分類に与える影響および効果について検討する上で重要となる属性ベクトル生成ツールについて述べる。

2. システム概要と背景

2.1 臨床症例クラスタリングシステム

基本システム構成を Fig.1 に示す。

システム (Fig.1) は、マップ生成のための SOM エンジンを中心に、情報視覚化エンジン (Visualization Engine)、データ検索・閲覧エンジン (Searching & Browsing Engine)、属性管理ツール (Attribute Manager) およびこれらを統合するユーザインタフェース部 (User Interface) から構成される。

属性管理ツールは、症例文書の全文データベース (Full-text Clinical Case Database) とリンクした文書属性およびインデキシング情報データベース (Attribute & Index Database) を制御する。

2.2 SOM と属性ベクトル

SOM は、トポロジカルマッピングを拡張した教師なし競合学習型ニューラルネットであり、入力層とマップ (出力) 層の 2 層構造をなす。また、データ間の特徴類似度による汎用的なクラスタリング能力を持つ。SOM を用いた文書情報検索システムとしては、WEBSOM^{[4],[5]} が知られている。

SOM モデルは、入力層では n 個、マップ層では 2 次元的に配列された m 個のニューロンからなる。入力層とマップ層の各ニューロンは全結合であり、それらの結合荷重は、 $m \times n$ 行列で表現される。

SOM の学習は、 n 次元入力ベクトル (= 属性ベクトル) を

マップ層に繰り返し投入し、結合荷重を更新することでなされる。

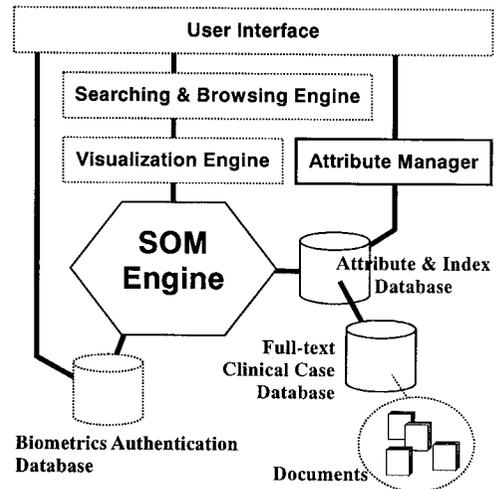


Fig.1 Basic System Construction

3. 属性ベクトル生成ツール

3.1 属性ベクトルの構成

クラスタリングマップを生成するために入力層に与えるデータ群は、クラスタリング対象の特徴属性と見なし得る任意の数量化データをベクトル表現したものであれば良い。

以下に、実際に我々が試みた属性値の例を示す。

- ① 症例文書中の医学間連用語の出現頻度
- ② 肝機能血液検査結果の数値データ
- ③ 胃内視鏡画像診断の医師所見に基づく数量化データ
- ④ 腹部超音波診断の医師所見に基づく数量化データ

3.2 属性値の重み変化

上記のような頻度データを扱う場合は、実際には、出現確率や共起確率を求める必要がある。それ以外の数値データや数量化データは種類の異なるものを混在させる場合、特定の属性値の影響を抑えるために、正規化・標準化を施す必要がある。単純な正規化の例としては、同一属性内の相対的な割合としてパーセント値を用いる方法がある。あるいは、特定のレンジ内に振

Development of Clustering System for Clinical Cases with Self-Organizing Maps

— Effect of Weight Change of Attribute Vectors for Document Clustering —

Kazuhiro NOTOMI¹⁾, Toshimitsu YAMAGUCHI¹⁾, Keiichi SAITO²⁾ and Tetsuo FUJIMOTO³⁾

1) Department of Information and Computer Sciences, Kanagawa Institute of Technology

2) Department of Business Management, University of East Asia

3) Department of Mechanical Engineering, Shibaura Institute of Technology

E-mail: notomi@ic.kanagawa-it.ac.jp

り分ける一次変換を考えることもできる。更に、属性の2値化を行なうことにより、論理型として扱うことも可能である。

多くの場合、各属性のレンジオーダーを揃えた方が良いが、ある属性値が他の属性値に対して独立でない場合、予想通りのクラスタリング結果とならない場合がある。

このように、SOM のクラスタリング性能を左右する最も重要なパラメータが属性ベクトルであり、この属性値のチューニングを簡便に行なうためのツールが必要不可欠となる。

3.3 属性ベクトル生成ツール

臨床症例クラスタリングにおける属性ベクトルは、元の症例報告文書に出現する医学関連用語(キーワード)の出現頻度が元になっている。そこで、以下のようなチューニングのための重み変化(①~④)を行なえる属性ベクトル生成ツールを試作した。ツールの動作画面例を Fig.2 に、生成された属性ベクトルの表示例を Fig.3 に示す。

- ①定数バイアス
- ②定数倍
- ③キーワードの文書間共有数倍
- ④文書内頻度順位倍

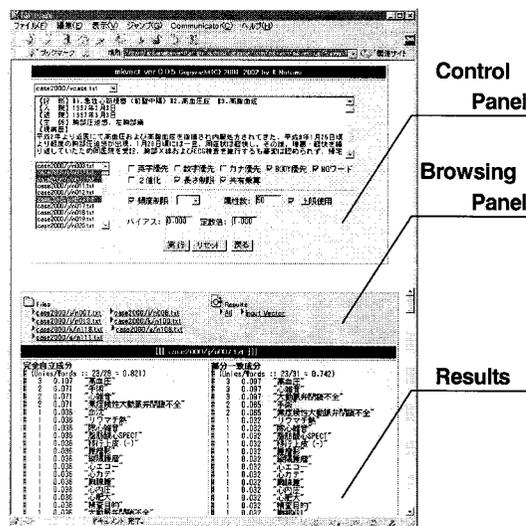


Fig.2 Attribute Vector Generating Tool

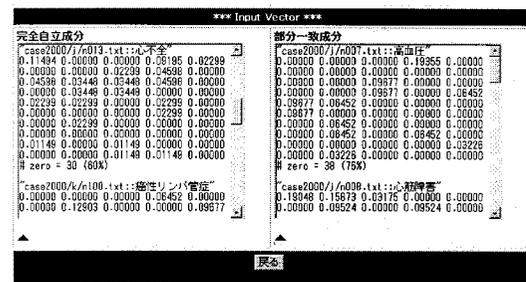


Fig.3 Sample of Attribute Vectors

本ツールを CGI ベースの Web アプリケーションプログラムとして UNIX(Linux)サーバ上に C++言語により実装した。パソコンなどから一般的な WWW ブラウザを用いることで、インターネットを経由して、サーバに接続し、利用することが可能である。

3.4 クラスタリングマップ

症例クラスタリングでは、症例文書から得られた情報を元に (Fig.1 参照)、入力ベクトルを生成し、これらに SOM アルゴリズムを適用してクラスタリングマップを生成することで、症例文書そのものを3疾患系 (①循環器系, ②消化器系, ③呼吸器系) 毎に分類するものである [1], [2], [3]。以下に例を示す。

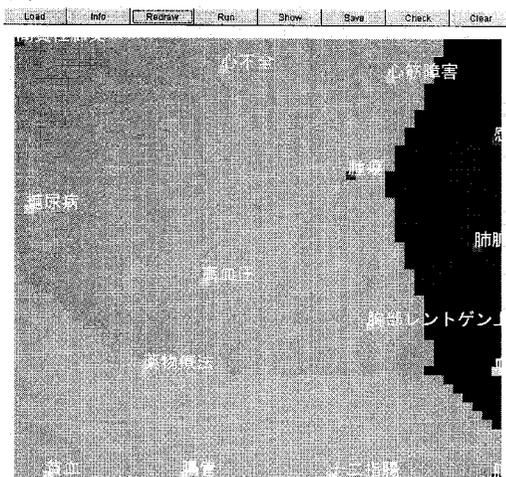


Fig.4 Sample of Clustering Map

4. まとめ

自己組織化マップ(SOM)を用いた症例クラスタリングシステムにおける属性ベクトルの重み変化による分類効果について述べた。

今後、定量的評価データを増やして行く予定である。

参考文献

- [1] 納富, 山口, 他: 自己組織化マップによる臨床症例クラスタリング-症例文書からの情報抽出ツールの汎用化と拡張-, 第 63 回情報処全大, 3U-01, (2001.03).
- [2] 納富, 山口, 他: 臨床症例データベース管理システムの構築-自己組織化マップによる情報の視覚化と検索インタフェース-, 第 62 回情報処全大, 6Q-01, (2001.03).
- [3] 納富, 岡本, 山口, 他: WWW による臨床症例検索システムの開発-自然言語処理と自己組織化マップを用いた疾患系分類-, 第 61 回情報処全大, 4R-06, (2000.10).
- [4] S.Kaski, K.Lagus, T.Honkela, T.Kohonen: Statical Aspects of the WEBSOM System in Organizing Document Collections, Computing Science and Statics, 29, pp.281-290, (1998).
- [5] K.Lagus and S.Kaski: Keyword selection method for characterizing text documents maps, ICANN '99, (1999).