

# 自己組織化マップによる臨床症例クラスタリング

3U-1

—症例文書からの情報抽出ツールの汎用化と拡張—

納富 一宏<sup>1)</sup> 山口 俊光<sup>1)</sup> 斎藤 恵一<sup>2)</sup> 藤本 哲男<sup>3)</sup>

<sup>1)</sup> 神奈川工科大学情報工学科

<sup>2)</sup> 東亜大学経営学部経営学科 <sup>3)</sup> 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) を用いた症例クラスタリングシステムの構築を行っている [1], [2] .

本稿では、医療支援、診断支援を目指し、医療情報向けの汎用クラスタリングシステムの構築手法について検討する。特に、自己組織化マップへの入力ベクトル生成に有効なツールの汎用化と拡張方法について述べる。

## 2. 臨床症例クラスタリング

### 2.1 自己組織化マップ (SOM) の概要

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 [3], [4] 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。

今、 $j$  番目の  $n$  次元入力ベクトルを  $x_j$ 、 $i$  番目の重みベクトルを  $w_i$  とすると、ベクトル間のユークリッド距離  $\|w_i - x_j\|$  を最小とする組を  $k$  とすると、SOM アルゴリズムによる重みベクトル  $w_k$  の更新は次式で示される。

$$w_k^{new} = w_k^{old} + \alpha (x_j - w_k^{old}) z_k \quad (1)$$

$\alpha$  は「学習率」と呼ばれ、学習回数  $t$  の単調減少関数である。 $Z$  は競合作用値であり、 $k$  に一致した場合のみ 1、それ以外では 0 を与える。

実際の学習では、 $k$  番目のニューロンの幾何学的近傍に

ついても式 (1) を適用し、重み更新を行なう。

### 2.2 症例クラスタリング

一般的な症例文書とは、A4 で数ページ (通常、全角 1,000~4,000 文字程度) の文書であり、記載医の所属・氏名、患者の氏名 (イニシャル)・年齢・性別・生年月日・主訴、診断、入退院の日付、担当医名 (受持、外来)、入院目的、現病歴、既往歴、家族歴、生活歴、入院時現症、入院時検査所見、退院時処方、問題点などが記載されている。

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] .

システム構成的にみると、マップ生成のための SOM エンジン、および視覚化インタフェースについては、臨床症例以外の対象についてもそのまま利用することができる。

そこで、システムの汎用化と拡張を、

- ① 入力ベクトルの構成
- ② SOM アルゴリズムの適用
- ③ 情報視覚化およびクラスタリング結果提示

という 3 つのフェーズについて検討する。

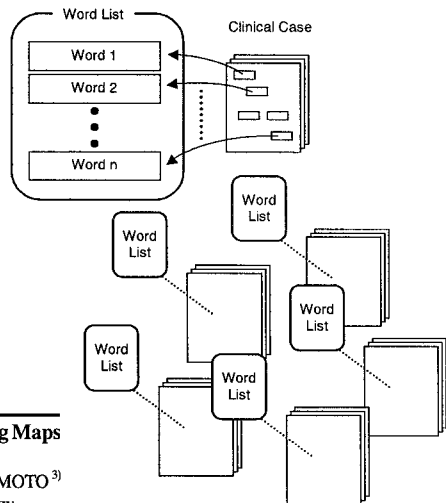


Fig.1 Keywords Extraction from Clinical Case

## Development of Clustering System for Clinical Cases with Self-Organizing Maps —Generalizing and Extending of Information Extraction Tool—

Kazuhiro NOTOMI<sup>1)</sup>, Toshimitsu YAMAGUCHI<sup>1)</sup>, Keiichi SAITO<sup>2)</sup> and Tetsuo FUJIMOTO<sup>3)</sup>

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

### 3. 汎用化と拡張

#### 3.1 入力ベクトルの構成

基本的に、入力層に与えるデータ群は、数量化された任意の属性値からなるベクトル表現であれば、クラスタリングマップを生成することは可能である (Fig.2 参照).

実際に我々が試みたクラスタリングマップ生成のための属性値の例としては以下のものがある.

- ① 症例文書中の医学関連用語の出現頻度
- 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

上記①のように、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、正規化を施す必要がある。

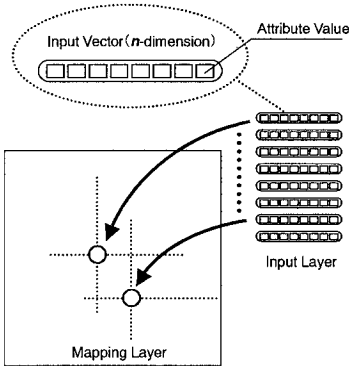


Fig.2 Data mapping of input vectors ( $n$ - dimension)

汎用データを扱う場合、クラスタリングを行なうべきデータと属性値との対応付けが重要である (Fig.3 参照).

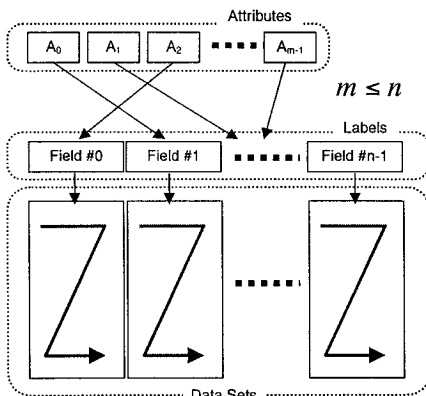


Fig.3 Attribute Selection of Data Sets

#### 3.2 SOM アルゴリズムの適用

SOM アルゴリズムの適用部分では、学習回数、マップサイズというパラメータの他に、学習率曲線、近傍判定曲線の

選択が可能である方が良い。これは、生成されたマップの良さを比べる上で必要な機能となる (Fig.4 参照).

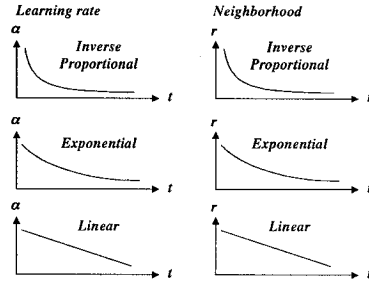


Fig.4 Variation of equation for SOM learning

#### 3.3 情報視覚化およびクラスタリング結果提示

マップ部分を階層構造化することでさまざまな情報提示が可能となる。クラスタリング領域のカラー化、属性強度のカラー化などが独立のレイヤーで操作できる方が良い (Fig.5 参照).

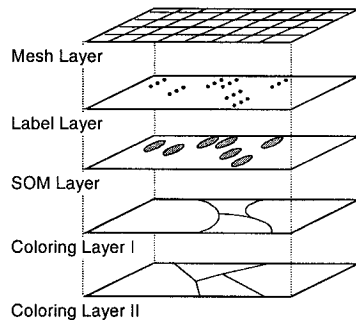


Fig.5 Display layers for SOM viewing

### 4. まとめ

自己組織化マップ (SOM) による臨床症例クラスタリングシステム構築における情報抽出ツールの汎用化と拡張について述べた。

今後の課題としては、医療情報全般を扱える汎用クラスタリングシステムについて更に検討を進める予定である。

#### 参考文献

- [1] 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 回情処全大, 6Q-01, (2001.03).
- [2] 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 回情処全大, 4R-06, (2000.10).
- [3] 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).
- [4] K.Lagus and S.Kaski : Keyword selection method for characterizing text documents maps, ICANN'99, (1999).