

納富 一宏¹⁾ 山口 俊光¹⁾ 斎藤 恵一²⁾ 藤本 哲男³⁾¹⁾ 神奈川工科大学情報工学科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

我々は以前から、病院等の医療関連施設における臨床症例報告書をWWWベースで管理する臨床症例データベース管理システムの構築を行なっており¹⁾、現在までに、自然言語処理および自己組織化マップ²⁾(Self-Organizing Map:以下, SOM)を用いた疾患系分類と症例検索への応用について検討してきた(2.1, 2.2 参照)。

臨床症例データの収集、蓄積、検索、閲覧には、信頼性、安全性、簡便性、操作性など多くのシステム要件が絡むため、これらの問題点を見極め、システムの設計を行なわなければならない。

そこで、本稿では、自己組織化マップによる症例文書のクラスタリング手法、視覚化手法、およびこれらを用いた症例検索インタフェースについて検討する。

2. 背景

2.1 症例報告書からのキーワード自動抽出

一般的な症例報告書(日本語)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、入院日の日付、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。

本システムでは、入力データとしての症例報告書からキーワードとなり得る医学関連用語を自動抽出する。

実際には、症例文書から検索キーワード候補を取得するために、①文字種別に着目した簡易形態素解析を行って一般用語を含む全ての自立語候補を抽出する。②次にこれらから系分類(後述)の際のノイズ低減を図るため、医学関連用語候補のみを選択する。この選択には、表記文字に関するヒューリスティックスを用いる。③選択された医学用語候補には重なりがあるため、ユニークな用語についてそ

れぞれの頻度(文書内異なり出現数)を求めた上で、頻度順リストを作成する。頻度情報はキーワードの重要度を示すパラメータとして用いる。

2.2 自己組織化マップ(SOM)による疾患系分類

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ワークであり、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^{2), 3)}が知られている。

症例文書を3疾患系(①循環器系、②消化器系、③呼吸器系)毎に分類するために、SOM アルゴリズムを適用する。SOM への入力ベクトルは、2.1 で述べた医学用語の頻度順リストから生成する。

実際には、系分類済みの症例サンプル(学習用データ)から、それぞれの頻度順リストを取得し、これらを合計した全体頻度順リストによる割合で正規化を施し、入力ベクトルを求める。

次に、この入力ベクトルで学習したマップ上に配置された同一疾患系に含まれるサンプル点の座標と、任意セルの座標から系毎の平均距離を求める。平均距離が最小となる系がその任意セルを占有すると考え、マップ全体の系領域を決定する。

以上から、任意の入力ベクトルが新たにマップに配置された時点で、その入力点が属する系を判定することができるようになる。

3. システム概要と動作例

3.1 情報の視覚化

SOM では、N 次元ベクトルで表現されたデータを 2 次元平面上へ写像することで、データ間の特徴類似度をプロット点間の距離という形で視覚化する。これにより、2 次元平面の近傍には、類似したデータが集まることになる。また、SOM アルゴリズムで学習に用いられる重み行列の値は、個々のベクトル要素である特徴属性の強さを表現していると考えられることができる。この属性の強さを輝度に対

Development of Database Management System for Clinical Cases

—Information Visualizing and Browsing Method with Self-Organizing Maps—

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

応させることで、属性による占有領域を色として視覚化することができる。

よって、症例文書を疾患系毎に分類する方法には、①疾患系への平均距離の最小値から判定する方法、②SOM学習の重み行列の値から判定する方法の2つが考えられる。

前者①は、分類済みの学習用データが必須となる。また、後者②は、入力ベクトルのどの属性を用いるかにより結果が変わってくる。これらの表示例を Fig.2, 3 に示す。共に、3 疾患系×15 文書=45 文書を学習用データとし、各文書からそれぞれ 30 個のキーワードを抽出し、5000 回の学習をさせた場合である。Fig.3 は、学習完了時点での重み行列のうち、分散値が高い属性値を RGB に対応させて表示したものである。

3.2 検索インターフェースへの応用

本システムの症例検索では、元来、自然言語インターフェースによる文書検索が可能である¹⁾が、今回はより直感的な視覚化インターフェースについて検討した。Fig.1 にインターフェースの画面例を示す。

検索では、ある症例に類似したいくつかの症例を選択することが目的となる。よって、SOM マップの領域をマウスドラッグで円形に範囲指定し、その領域に含まれるプロット点をマップから読み出す方式を採用した。

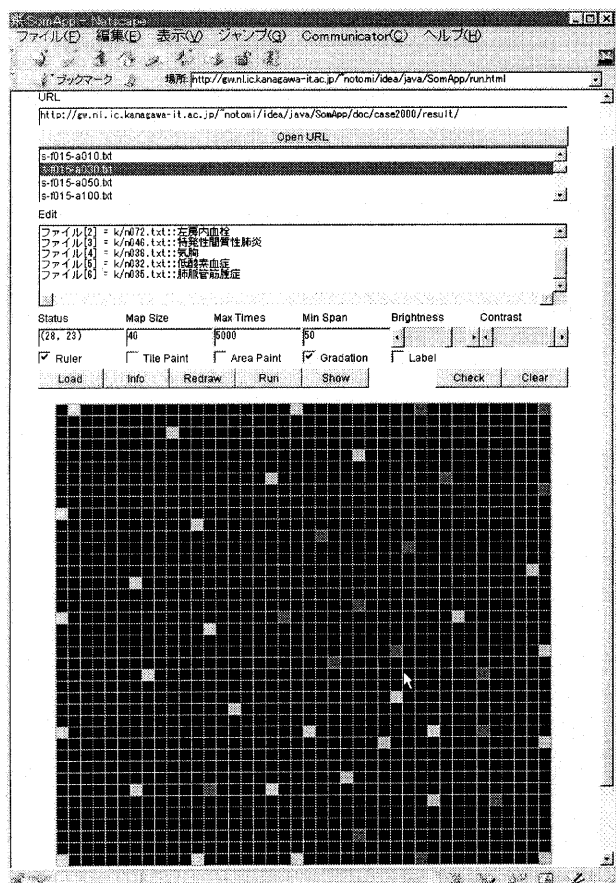


Fig.1 "SomApp" : Visualizing & Browsing Interface with SOM

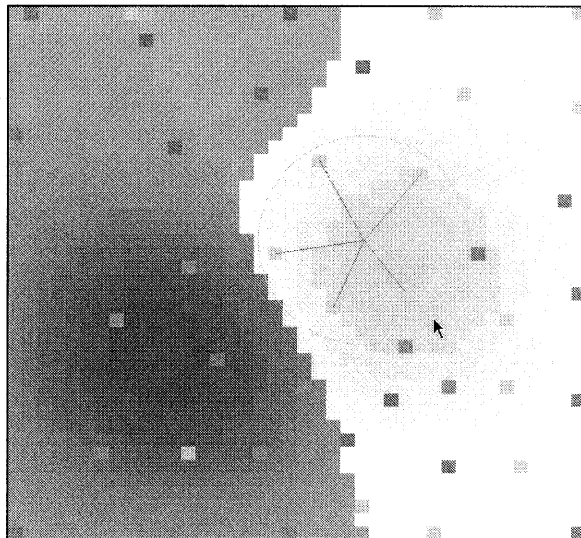


Fig.2 A SOM-map colored by the minimum distance

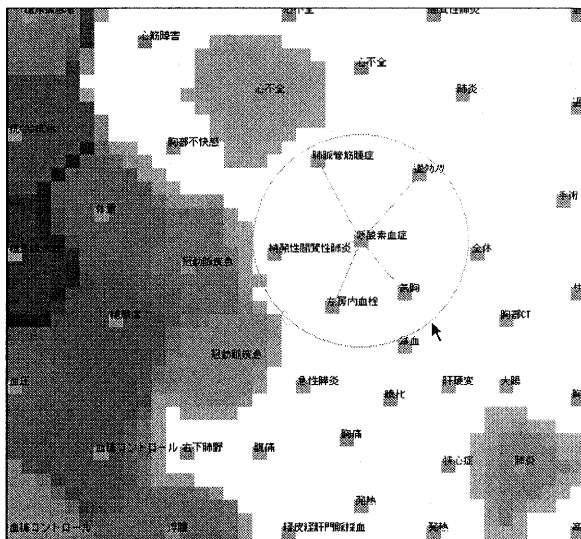


Fig.3 A SOM-map colored by weight values

4. まとめ

自然言語処理と自己組織化マップ(SOM)を用いた疾患系分類における情報の視覚化、および症例検索インターフェースへの応用について述べた。

現在の実装では、円形領域の範囲指定により類似症例を選択・閲覧することができる。また、複数の円形領域のAND 検索、OR 検索などへの対応も可能である。

参考文献

- [1] 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).
- [2] 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).
- [3] K.Lagus and S.Kaski : Keyword selection method for characterizing text documents maps, ICANN '99, (1999).