

納富 一宏¹⁾ 岡本 雅幸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

病院をはじめとする各種医療関連施設や医局などで行なわれている症例検討会で提示される症例報告書の多くは、最初からコンピュータ上で作成された電子化文書であるが、実際にはプリントアウトされた状態で配布され、個人レベルで管理されているのが現状である。これら過去の症例を情報リソースとして活用することが、新たな症例の病態を把握、検討するために非常に重要である。

また、近年では、インターネットの普及に伴い、医療現場にもコンピュータネットワークがすでに導入されている場合が多く、電子メールや Web による情報検索のために、日常的に広く利用されていると考えられる。

こうした状況の中、全国的な地域をカバーするための優位性を有するインターネットを利用し、簡便に臨床症例文書の検索・閲覧が可能なシステムの開発は、大いに期待されるであろう。

しかしながら、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う必要がある。

我々は以前から WWW による臨床症例データベース検索システムの構築を行なっている^[1]。本稿では、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述べる。また、先のシステム要件のうち、簡便性、操作性に着目し、データベース検索部における自然言語インタフェースを検討する。

2. システム概要と実現方法

2.1 自然言語による検索

類似症例を過去の報告データから検索する場合、従来型のデータベース検索では、キーワード指定や条件検索のための論理式の指定が必要となるため、ユーザは適切なキーワードを考慮した入力操作が要求される。

そこで、自然言語処理技術を応用し、キーワード指定の自動化を目標とした。このことで、本システムの症例データベースの検索では、①新規症例文書による類似検索、②任意の自然言語文章による検索、③任意キーワードによる検索、の3つの形式に対応することができる。

2.2 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)には、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おり、A4 で数ページ(通常、全角 1,000~4,000 文字程度)の文書である。

これらの文書の記載内容から、3疾患系(①循環器系、②消化器系、③呼吸器系)への分類を目的に、医学関連用語を選択・取得する。

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。

以上の操作が自然言語処理(Natural Language Processing: 以下, NLP)部分となる。

2.3 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。

症例文書を疾患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、2.2 で述べた医学用語の頻度順リストから生成する。

Development of DBMS for Clinical Cases on WWW

—Data Clustering Method with Natural Language Processing and Self-Organizing Maps—

Kazuhiro NOTOMI¹⁾, Noriyuki OKAMOTO¹⁾, 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

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。この操作手順が SOM 計算部分となる。

システムによる処理を Fig.1 にまとめる。

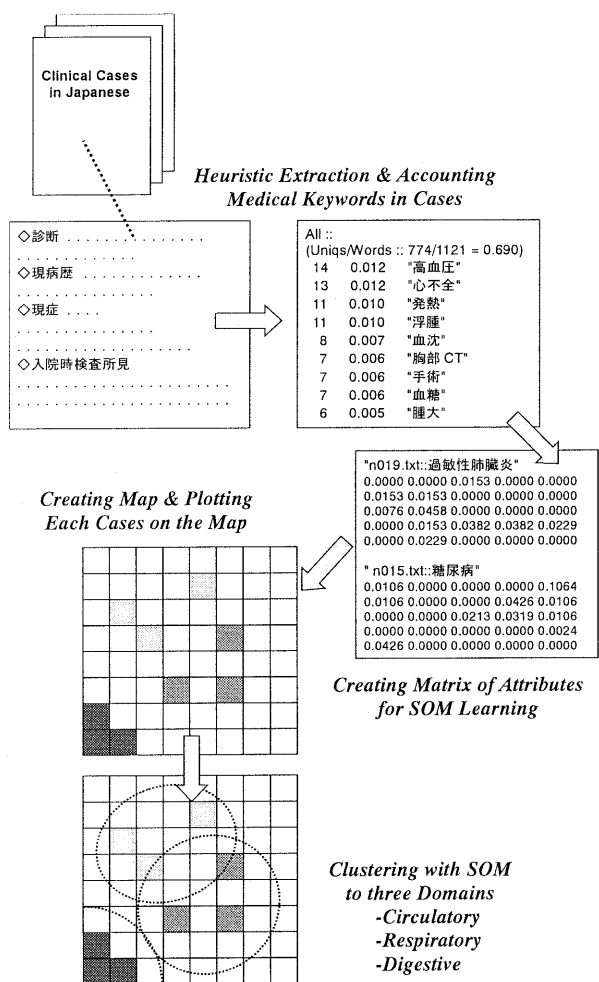


Fig.1 Clustering process with NLP and SOM

3. 評価実験

SOM による系分類の性能評価を目的として、以下の評価実験を行なった。

3.1 方法

臨床症例報告書 88 件をあらかじめ人手により3系(循環器系:C, 消化器系:R, 呼吸器系:D)に分類する。さら

に、それぞれの系に属するデータを学習用、および評価用に2分する。学習用データが(C, R, D)=(10, 10, 10), (15, 15, 15), (20, 20, 20) のそれぞれの場合について、セル数 20×20, 50×50, 80×80, 100×100 のマップを計算し、系領域を求める。次に評価用データをマッピングして先の領域から系を判定し、最初分類と比較して、正解率を求める。なお、入力ベクトルの属性数は、30, 50, 100 の場合について調べる。

3.2 結果

各学習回数における属性数 30, 50, 100 についての、マップサイズと正解率との関係を Fig.2 に示す。

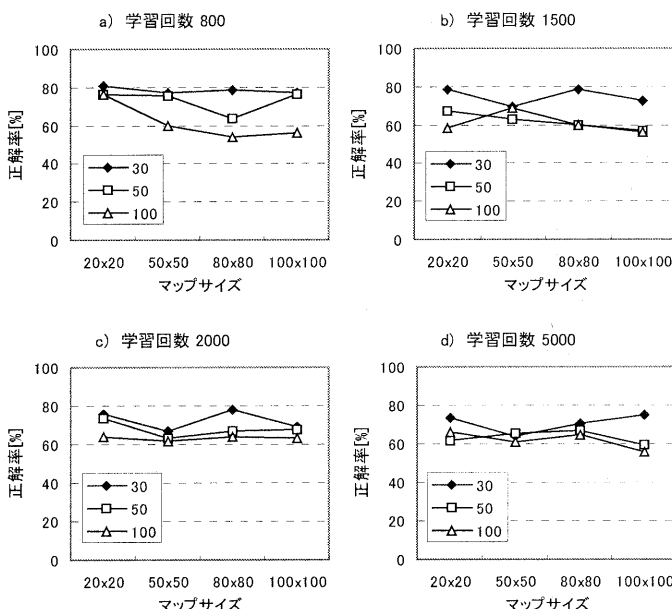


Fig.2 Results of evaluation on each map sizes and Learning Times

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ついて述べた。評価実験から、文書属性数 30 の入力ベクトルによるマップを用いた系分類正解率平均が 73.3%であり、本手法が疾患系分類に有効であることが分かった。

現在、データ検索時の直感的な情報視覚化インターフェースとして系分類時のマップを用いる場合について検討を進めている。

参考文献

[1] 納富, 斎藤, 他: WWW による臨床症例データベース検索システムの構築—自然言語処理による検索機能の実現—, BMFSA 第 11 回年次大会講演論文集, (1998.11).

[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).