

自己組織化マップによる 日本語文書の分類と全文データベースへの応用

納富 一宏¹・岡本 雅幸²・山口 俊光³・斎藤 恵一⁴・藤本 哲男⁵

¹ 情報工学科 ² 博士前期課程 情報工学専攻 ³ 情報工学科 学部3年

⁴ 東亜大学経営学部経営学科 ⁵ 芝浦工業大学工学部機械工学科

A Clustering Method of Japanese Documents by Self-Organizing Maps for Full Text Database

Kazuhiro NOTOMI¹⁾, Noriyuki OKAMOTO²⁾, Toshimitsu YAMAGUCHI³⁾,
Keiichi SAITO⁴⁾, Tetsuo FUJIMOTO⁵⁾

Abstract

In this article, we propose a document clustering method with Natural Language Processing and Self-Organizing Maps (SOM) on our Web based DBMS (Data Base Management System) for clinical cases. The system is implemented in Java language on the Internet server, and all documents of clinical cases in the database are written in Japanese. We tested the system ability to evaluate for the correct clustering of clinical cases, that each document has 30 attributes in an input-vector for SOM-learning. For the result of this ability test, it shows that 73.9% of clustering datasets are correct answers. Thus, it seems that our method is available to search the similar cases in the database.

Key Words : Self-Organizing Map, Neural Network, Document Clustering, Natural Language Processing, Full Text Database, Clinical Case

1. はしがき

病院をはじめとする各種医療関連施設や医局などでは、臨床医(専門医)が日常扱う症例とは異なる希有な症例や複雑な症例に出会うと、各施設で定期的に開かれる症例検討会などで報告、検討がなされる。また、そのうちの特別なものは専門の学会誌や論文誌に掲載される。これらのことは、新規症例の病態の把握や問題点の考察を行なう上で大変重要であり、近年では、臨床症例データベースの構築およびそのネットワーク化が進められてきている^[8]。

症例検討会で提示される症例報告は、日本語で記述されたやや定型的な文書である。一般には「カルテ」に類する文書であると考えて良い(付録参照)。

この特殊な日本語文書から全文データベースを構築する場合、あらかじめ疾患系毎の分類を行なうことで、データベースの分散化が期待できる。このことは、システムの検索性向上を目指す上で重要であると考えられる。

我々は以前からこのテーマに取り組み、WWW による臨床症例データベース検索システムの検討を行なってきた^[1]。

具体的には、臨床症例データの収集、蓄積、検索、閲覧に

おける信頼性、安全性、簡便性、操作性など多くのシステム要件を見据えたシステム設計と、ベースとなるシステムの実装を行ない、システムリファインのための疾患系分類手法として、ファジィ測度による症例分類^[1]、^[2]、アソシエイトロンによる疾患系分類^[3]について検討してきた。

そこで本稿では、臨床症例報告書を日本語文書として捉え、自然言語処理(Natural Language Processing:以下、NLP)および自己組織化マップ^[2](Self-Organizing Map:以下、SOM)を用いた臨床症例文書の分類(clustering)手法を提案し、本手法によるクラスタリング性能の評価実験結果を示すと共に、全文データベースへの応用としての症例検索について述べる。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について検討する。

2. 手法の提案と実装システムの概要

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(AND, OR, NOT, etc.)の指定が必要であるため、ユーザには適切なキーワードを考慮した入力操作が要求される。特に、データベースの規模が大きくなればなるほど、検索対象の絞り込みを行なう操作が重要となるため、ユーザへの負担も増すことになる。

こうした情報検索の際のユーザへの負担を低減するために、システムに自然言語インタフェースを実装することで、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可能になった。

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呼吸器系、③消化器系)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高速簡易形態素解析(JFK 分解^{[6], [7]})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次にこれらから疾患系分類(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(高頻度優先)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^{[4], [5]}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(高頻度優先)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順位により入力ベクトルを求める。

以下、この手順を定式化する。

入力となる症例文書から m 個の異なり単語 w_i が抽出され、その単語頻度が n_i であるとする。単語頻度取得関数を $Freq$ として、頻度順リスト L は次式で定義される。

$$L = [w_1, w_2, \dots, w_m] \quad \dots (2.1)$$

$$n_i = Freq(w_i), \quad n_1 \geq n_2 \geq \dots \geq n_m \quad \dots (2.2)$$

また、頻度順リストの異なり単語頻度を再集計する関数を $Uniq$ とすると、 N 文書からなる全体頻度順リスト L_{all} は、次式のようになる。

$$L_{all} = Uniq\left(\bigcup_{k=1}^N L_k\right) \quad \dots (2.3)$$

$$= [w_1^{all}, w_2^{all}, \dots, w_n^{all}], \quad w_i^{all} \neq w_j^{all}$$

今、 L_{all} に含まれる単語 w_i が頻度順リスト L に含まれる単語 w_r に一致した場合の頻度 n_r を求める関数 $SearchKeyOrder$ を次式で定義する。

$$n_r = SearchKeyOrder(w_i, L), \quad (w_i \in L_{all})$$

$$= \begin{cases} Freq(w_r) & \text{if } (w_i = w_r \in L) \\ 0 & \text{if } (w_i \notin L) \end{cases} \quad \dots (2.4)$$

SOM 学習のための入力ベクトル x は、次式となる。

$$x = (x_1, x_2, \dots, x_n), \quad x_i = \frac{n_r}{\sum_{k=1}^m n_k} \quad \dots (2.5)$$

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 s_i と、任意セルの座標 s とから系毎の平均ユークリッド距離 d を求める。平均距離が最小となる疾患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。

疾患系 Z (=C:循環器系 or R:呼吸器系 or D:消化器系)のサンプル点の個数が n 個である場合、距離関数 d_z を次式で定義する。

$$d_z(s) = \frac{1}{n} \sum_{i=1}^n \|s - s_i\| \quad \dots (2.6)$$

領域決定関数 $Cluster$ は次式となる。

$$Cluster(s) = \begin{cases} "C" & \text{if } (d_{\min} = d_C(s)) \\ "R" & \text{if } (d_{\min} = d_R(s)) \\ "D" & \text{if } (d_{\min} = d_D(s)) \end{cases} \quad \dots (2.7)$$

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。

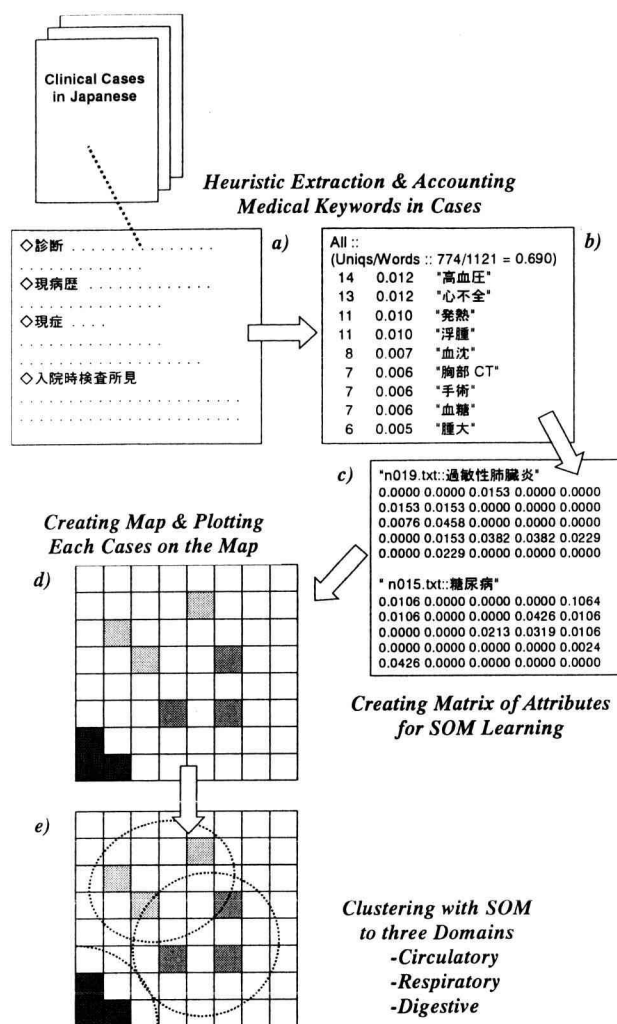


Fig.1 Clustering process with NLP and SOM

a) Clinical cases in Japanese, b) List of unique medical keywords sorted with its own frequency, c) Matrix of attributes to input, d) Self-organizing feature map for document clustering, e) 3-domain clustering

2.4 実装システム概要

システムは Java 言語により実装されており、先に示した通り、NLP 部 (クラス数:5, ソース行数:680), および SOM 部 (クラス数:5, ソース行数:1252) からなる。

マップ生成の動作画面例を Fig.2 に示す。

Fig.2 は、一般的なブラウザから Java アプレットを起動した画面 (計算条件: マップサイズ 80×80, 学習回数 800, 属性数 30, 入力ベクトル数 10) である。

図中、ウィンドウのクライアント領域上半分に示された各種コントロールが、計算条件を設定するインターフェースになっており、画面下半分に計算結果としてマップが表示される。

SOM 学習によりマッピングされた症例文書は高輝度のピクセルとして表示され、マップ全体について、3疾患系毎に色分けされた領域を確認することができる。この領域がクラスタリング領域となる。クラスタリング領域はグラデーション表示されており、同一疾患系に含まれる症例文書の配置点の重心から辺

縁部に向けて輝度が徐々に下がっていく。この減少率は、インタフェースの Brightness 値および Contrast 値で指定することができる。

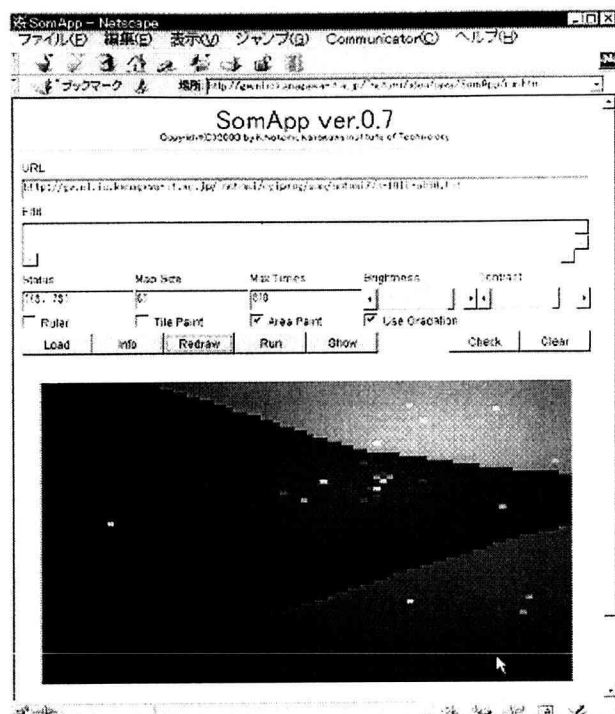


Fig.2 Screenshot of Document Clustering with SOM

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, 呼吸器系: R, 消化器系: D) に分類する。さらに、それぞれの系に属するデータをデータサイズ順に整列後、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。

評価条件とデータ数の関係を Table.1 にまとめる。なお、表中の三つ組みは、それぞれ C, R, D の症例文書数を表す。

Table.1 Conditions of Evaluation

データ数 (C, R, D)	学習用	(10, 10, 10)	(15, 15, 15)	(20, 20, 20)
	評価用	(26, 15, 17)	(21, 10, 12)	(16, 05, 07)
入力ベクトル属性数		30	50	100
マップサイズ		20x20	50x50	80x80
学習回数		800	1500	2000

Table.1の評価条件は、データ数条件が3, 属性数条件が3, マップサイズ条件が 4, 学習回数条件が 4 パターンあり, 総計 144(=3×3×4×4)回の試行を意味する。

なお, 実際の症例報告文書の例(仮症例)を付録に示す。

3.2 結果

学習データ数 10, 15, 20, 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 Table.2 に示す。また, 学習回数毎の平均正解率の変化をグラフとして Fig.3 に示す。

Table.2 について補足する。

マップサイズは, 生成するマップの解像度を示すパラメータであるが, 同時に疾患系の分類精度基準を示すパラメータであるとも言える。症例文書のクラスタリングを行なう場合, マップサイズを大きくすることは, 精度基準を引上げることを意味し, 結果としてよりシビアな評価を要求する。今回の実験では, 正解率マップ平均の変動範囲は, マップサイズ(セル数)が 20×20 のとき 57.7~81.0[%]であり, 100×100 のとき 55.7~77.7[%]となっている。変動範囲の下限値は, 4 種類のマップサイズでほぼ均一となり, 平均で 57.2[%]($\sigma = 1.0$ [%])となる。上限値は平均で 77.4[%]($\sigma = 2.6$ [%])である。

属性数は, 本来, 分類データに含まれる特徴や性質の個数であるが, 今回の場合は, 各症例報告に含まれる医学関連用語候補の出現確率の個数となっている。2章で述べた通り, 頻度順リスト(高頻度優先)を利用することから, 属性数が少ない場合, 高頻度の用語のみから入力ベクトルが構成されることになり, 結果として都合がよい。反面, 属性数が多くなると, ノイズとなる用語の割合が増える。実験では, 正解率属性平均の変動範囲は, 属性数 30 のとき 49.8~83.5[%]であり, 属性数 100 のとき 55.8~68.3[%]となっている。変動範囲の下限値は, 3 種類の属性値の中で 100×100 のときがやや高く, 平均で 51.8[%]($\sigma = 3.5$ [%])となる。上限値は平均で 77.2[%]($\sigma = 7.9$ [%])である。

学習回数は, 計算精度を高めるパラメータであるが, 計算時間とのトレードオフを常に意識する必要がある。今回の場合, 5000 回までの評価しか行なわなかったため, それぞれの学習回数の正解率への影響に関しては, 特別な差は見られなかった。この点については5章で考察する。

次に, Fig.3 について補足する。

各グラフ a)~d)を見ると, 学習回数が増えるほど, 各条件の正解率の変動範囲は小さくなっていくことが分かる。実際, 学習回数 800 回のとき標準偏差 $\sigma = 9.7$ [%]となり, 1500, 2000, 5000 回で, それぞれ $\sigma = 7.9$ [%], $\sigma = 5.3$ [%], $\sigma = 4.6$ [%]となる。また, 変動範囲の下限値は, 学習回数が増えるほど, 53.7, 56.0, 61.7, 59.3[%]と変化し, 上昇傾向にあるのに対し, 逆に上限値は, 80.7, 78.3, 77.7, 75.0[%]と変化し, 下降傾向にあることが分かる。平均値は 70.8, 65.6, 67.5, 66.9[%]と変化し, やや下降傾向を示す。

Table.2 Results of Evaluation

a) 学習データ数:(C, R, D)=(10, 10, 10)

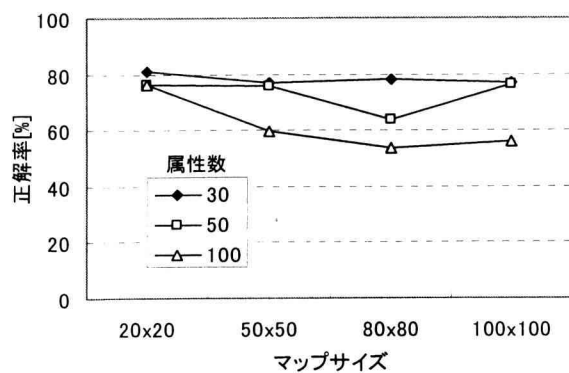
学習回数	マップサイズ毎の正解率 [%]				
800	20x20	50x50	80x80	100x100	属性平均
属性数	30	82	73	67	84
	50	76	76	78	89
	100	73	56	51	60
マップ平均	77.0	68.3	65.3	77.7	72.1
1500	20x20	50x50	80x80	100x100	属性平均
属性数	30	87	62	71	67
	50	71	64	64	60
	100	58	56	67	64
マップ平均	72.0	60.7	67.3	63.7	65.9
2000	20x20	50x50	80x80	100x100	属性平均
属性数	30	84	62	89	69
	50	73	69	73	62
	100	60	62	64	69
マップ平均	72.3	64.3	75.3	66.7	69.7
5000	20x20	50x50	80x80	100x100	属性平均
属性数	30	76	62	62	71
	50	71	69	69	64
	100	58	56	80	49
マップ平均	68.3	62.3	70.3	61.3	65.6

b) 学習データ数:(C, R, D)=(15, 15, 15)

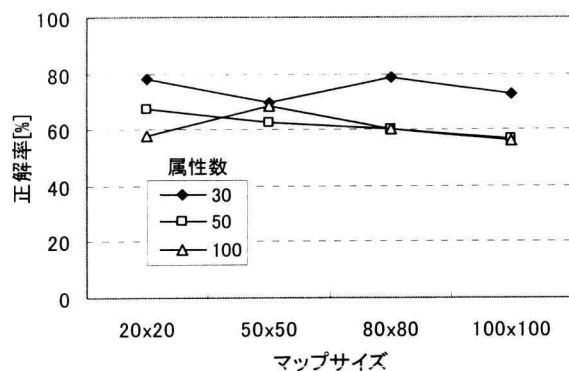
学習回数	マップサイズ毎の正解率 [%]				
800	20x20	50x50	80x80	100x100	属性平均
属性数	30	80	70	87	87
	50	80	70	60	80
	100	83	63	63	60
マップ平均	81.0	67.7	70.0	75.7	73.6
1500	20x20	50x50	80x80	100x100	属性平均
属性数	30	80	87	97	70
	50	77	70	63	70
	100	63	63	60	57
マップ平均	73.3	73.3	73.3	65.7	71.4
2000	20x20	50x50	80x80	100x100	属性平均
属性数	30	70	77	77	77
	50	80	67	67	73
	100	63	63	60	60
マップ平均	71.0	69.0	68.0	70.0	69.5
5000	20x20	50x50	80x80	100x100	属性平均
属性数	30	77	70	63	87
	50	60	67	70	67
	100	53	53	60	57
マップ平均	63.3	63.3	64.3	70.3	65.3

c) 学習データ数:(C, R, D)=(20, 20, 20)

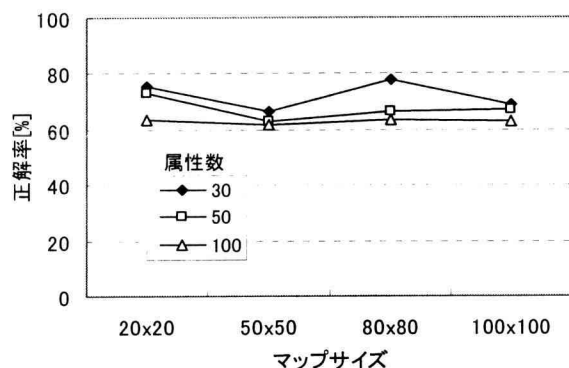
学習回数	マップサイズ毎の正解率 [%]				
800	20x20	50x50	80x80	100x100	属性平均
属性数	30	80	87	80	60
	50	73	80	53	60
	100	73	60	47	47
マップ平均	75.3	75.7	60.0	55.7	66.7
1500	20x20	50x50	80x80	100x100	属性平均
属性数	30	67	60	67	80
	50	53	53	53	40
	100	53	87	53	47
マップ平均	57.7	66.7	57.7	55.7	59.4
2000	20x20	50x50	80x80	100x100	属性平均
属性数	30	73	60	67	60
	50	67	53	60	67
	100	67	60	67	60
マップ平均	69.0	57.7	64.7	62.3	63.4
5000	20x20	50x50	80x80	100x100	属性平均
属性数	30	67	60	87	67
	50	53	60	60	47
	100	87	73	53	60
マップ平均	69.0	64.3	66.7	58.0	64.5



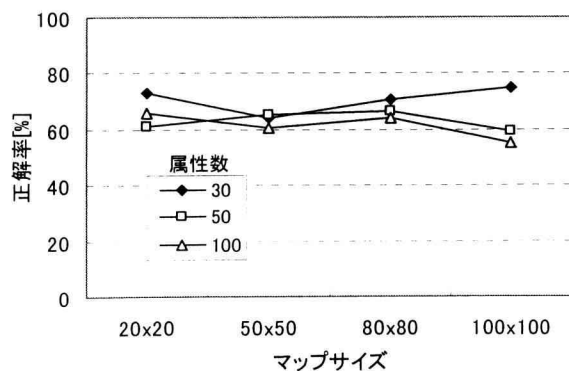
a) 学習回数 800



b) 学習回数 1500



c) 学習回数 2000



d) 学習回数 5000

Fig.3 Results of Evaluation on each Map Sizes and Learning Times

4. 考察

4.1 属性数と学習データ数について

学習データ数 10, 15, 20 の総平均を Table.3 に示す。

SOM 学習は、結合重みの初期値として乱数を用いる。従って、学習毎に生成されるマップは異なったものとなる。評価実験では条件の組み合わせが144通りにのぼり、各条件で1試行ずつのデータしか得られなかった。この点を考慮しなければならない。

Table.3 Total Average of the Results

総平均(学習データ平均)

学習回数		マップサイズ毎の正解率[%]				属性平均
800		20x20	50x50	80x80	100x100	
属性数	30	80.7	76.7	78.0	77.0	78.1
	50	76.3	75.3	63.7	76.3	72.9
	100	76.3	59.7	53.7	55.7	61.3
マップ平均		77.8	70.6	65.1	69.7	70.8
1500		20x20	50x50	80x80	100x100	属性平均
属性数	30	78.0	69.7	78.3	72.3	
	50	67.0	62.3	60.0	56.7	61.5
	100	58.0	68.7	60.0	56.0	60.7
マップ平均		67.7	66.9	66.1	61.7	65.6
2000		20x20	50x50	80x80	100x100	属性平均
属性数	30	75.7	66.3	77.7	68.7	
	50	73.3	63.0	66.7	67.3	67.6
	100	63.3	61.7	63.7	63.0	62.9
マップ平均		70.8	63.7	69.3	66.3	67.5
5000		20x20	50x50	80x80	100x100	属性平均
属性数	30	73.3	64.0	70.7	75.0	
	50	61.3	65.3	66.3	59.3	63.1
	100	67.3	64.7	68.5	67.2	66.9
マップ平均		67.3	64.7	68.5	67.2	66.9

評価実験の結果から、属性数 30 の場合が最も成績が良く、全マップサイズにおける属性平均で見ると、学習回数平均で正解率は73.9[%]となる。他の属性数に対する正解率の変化を Fig.4 に示す。

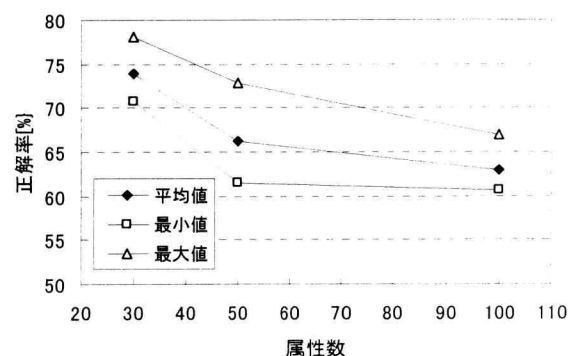


Fig.4 Match Rate on each Attributes

結果でも述べたが、単に正解率が高いだけでは分類(クラスタリング)性能を判断することが難しい。Fig.4 に見られる属性数は、1文書あたりの異なりキーワード(特徴)の個数であり、30個で少ないわけではない。Web の検索エンジンでも、キーワード指定が 30 以上となるのはまれである。逆に、100 個もの

属性を扱うためには、元の文書サイズがある規模以上でなければならないことが分かる。今回の実験でも、属性数 100 の場合に、抽出キーワード数が 100 に満たない場合もあった。よって、本手法の実用化では、属性数が 30 以下に設定することが望ましいものと結論できる。

次に学習データ数であるが、本稿で述べている頻度順リスト(高頻度優先)を利用する場合、1疾患系あたりの学習データ数 15(全体の 51%)のときが最も成績が良く、平均で 70.0[%]となる。データ数 10(全体の 34%)、20(全体の 68%)のときがそれぞれ 68.3、63.5[%]であることから、データ数がある値以上となると、ノイズの影響が増加するものと考えられる。このことから、分類したい文書データサイズを考慮し、全体の文書数から判断して、50%以下となるようにすればよいことが判る。

4.2 マップサイズと学習回数について

前節から正解率の高い結果が得られる条件として、属性数 30、学習データ数(1疾患あたり)15 のデータに注目すると、マップサイズ毎の正解率は、20×20、50×50、80×80、100×100 で、それぞれ 76.8、76.0、81.0、80.3 [%]となる(Fig.5)。

また、同じ条件で、学習回数毎の正解率は、800、1500、2000、5000 で、それぞれ 81.0、83.5、75.3、74.3 [%]となる(Fig.6)。

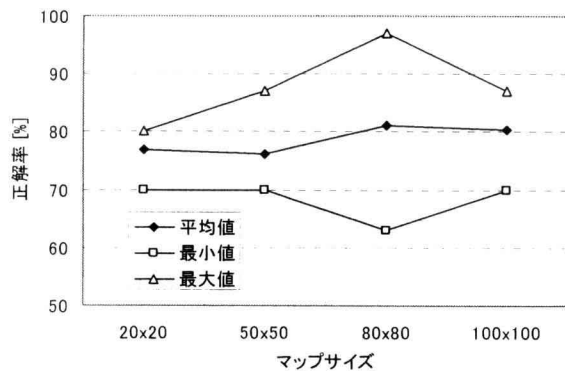


Fig.5 Match Rate on each Map Sizes

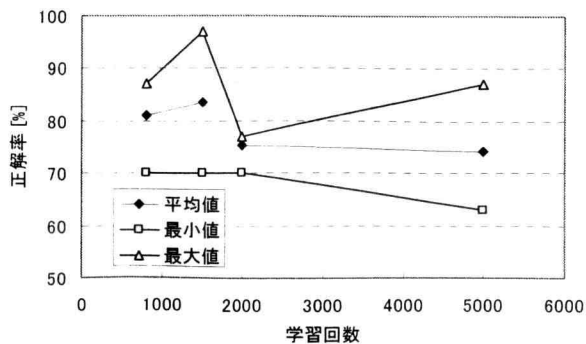


Fig.6 Match Rate on each Learning Times

これらのグラフから、総合的な判断をすると、マップサイズについては、50×50 以上、80×80 以下の範囲で決定すべきで

あることが判る。また、学習回数については、1500～2000 回の範囲が妥当であると言える。100×100 以上のマップや、5000 回以上の学習については、データが得られていないため不明である。

評価実験における最適な SOM 学習条件を Table.4 にまとめる。

Table.4 Optimum Conditions

学習データ数	(C, R, D) = (15, 15, 15)
属性数	30
マップサイズ	50x50 ~ 80x80
学習回数	1500 ~ 2000

上記条件に合致する試行は 4 試行あり、正解率の平均値は、84.5[%]となる。参考までに、マップサイズ範囲を 50×50～100×100、学習回数範囲を 1500～5000 に拡大すると、試行数は 9 となり、その正解率平均は 78.3[%]となる。また、属性数 30 のみを条件とした場合、試行数は 48 となり、正解率平均は 73.9[%]となる。

4.3 単語重複率について

疾患系毎の症例文書の統計値を Table4 にまとめる。

Table.5 Statistics of Documents

		学習データ数		
		10	15	20
a) 循環器系				
平均テキストサイズ [KB]		9.04	7.86	7.00
平均抽出単語数	異なり	54.60	49.00	44.40
	重なり	71.00	63.00	56.20
平均単語重複率		1.30	1.29	1.27

		学習データ数		
		10	15	20
b) 消化器系				
平均テキストサイズ [KB]		3.51	2.96	2.52
平均抽出単語数	異なり	35.00	28.87	23.25
	重なり	44.80	38.13	31.05
平均単語重複率		1.28	1.32	1.34

		学習データ数		
		10	15	20
c) 呼吸器系				
平均テキストサイズ [KB]		5.50	4.49	3.72
平均抽出単語数	異なり	51.00	47.40	40.40
	重なり	63.50	57.80	49.70
平均単語重複率		1.25	1.22	1.23

Table.4 において、単語重複率とは、NLP 部分において自動抽出された医学関連用語候補の重なり数と異なり数との割合である。従って、この値の最小値は 1 であり、このとき単語の重複はない。逆に、この値が大きいほど、同一単語の使用頻度が高いと見なせる。

本稿で述べた SOM 学習のための入力ベクトル値は、頻度順リスト(高頻度優先)から生成されるため、単語重複率が重要なファクターとなる。各単語頻度が一定であると、単語間の重みが均一となり、特徴点が失われるため、クラスタリング(分類)性能が低下する恐れがある。故に、分類性能を一定レベルより高めるためには、①ある程度の文書サイズを有し、②単語頻度の分散値が大きく、③単語重複率が高い、文書集合を学習

データとして選択する必要があるものと思われる。

5. むすび

自己組織化マップ(SOM)を用いた日本語文書の分類手法について述べた。本手法では、自然言語処理(NLP)を用いて文書データから SOM 学習のための入力ベクトルを生成した。また、本手法の応用例として、臨床症例データベースにおける疾患系分類の評価実験(計 144 試行)を行なった。

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(計 48 試行中)が 73.9%であり、本手法が日本語文書の分類に有効であることが確認された。

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

参考文献

- [1] 納富, 斎藤, 藤本: WWW による臨床症例データベース検索システムの構築, バイオメディカル・ファジィ・システム学会誌, Vol.1, No.1, pp.35-45(1999).
- [2] 斎藤, 納富, 藤本: WWW による臨床症例データベース検索システムの構築(第2報)ーファジィ測度論による症例分類ー, 情報処理学会第 58 回全国大会講演論文集, 5G-03 (1999).
- [3] 納富, 斎藤, 藤本: 自然言語処理とアソシアトロンを用いた疾患系分類, 情報処理学会第 58 回全国大会講演論文集, 5G-03 (1999).
- [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).
- [6] 納富, 他: 「日本語文書校正支援システムにおける高速統計解析手法」, 神奈川工科大学研究報告, B理工学編, 第 20 号, pp.165-175, (1997).
- [7] 納富: 日本語文書校正支援ツール HSP の開発, 情報処理学会デジタルドキュメント研究会報告, (1997).
- [8] 開原成允, 大江和彦, 長瀬淑子, 木内貴弘, 桜井恒太郎, 根岸正光, 小山照夫, 志津田嘉康, 熊淵智行, 天野善雄, 裏田和夫, 野添篤毅: 新しい考え方の医学用語シソーラスを用いた臨床症例データベースの作成, 医療情報学会雑誌, 1993, 13, 221-228..

付 録 一症例報告文書の例一

【患 者 名】A.A 60 歳 男性

【生年月日】1937 年 1 月 1 日

【診 断】#1.急性心筋梗塞(前壁中隔)
#2.高血圧症 #3.高脂血症

【入 院】1997 年 2 月 3 日

【退 院】1997 年 3 月 3 日

【主 訴】胸部圧迫感, 左胸部痛

【現 病 歴】

平成 2 年より近医にて高血圧および高脂血症を指摘され内服処方されてきた。平成 9 年 1 月 25 日頃より軽度の胸部圧迫感が出現。1 月 28 日頃には一旦、同症状は軽快し、その後、増悪・軽快を繰り返していたため同医院を受診。胸部 X 線および ECG 検査を施行するも著変は認められず、帰宅し経過観察とされた。2 月 3 日の夕食後(20:00 頃)、強い左胸部痛が出現し、我慢できなくなったため同医院を再度受診。ECG では異常を認められなかったが、虚血性心疾患も否定しきれないためフランドルテープとニトログリセリンを処方された。しかし帰宅後も同症状はさらに増悪したため、同日 22:00 救急車にて当院救急センターを受診。ECG 上 V1~V5 で ST 上昇、心臓超音波検査(心エコー)上で前壁中隔梗塞の hypokinesis を認めたため、同部位の梗塞が疑われ、当科緊急入院となった。

【既 往 歴】1971 年;肺炎,

1990 年より;高血圧症, 高脂血症(通院中)

【家 族 歴】父;心筋梗塞, 糖尿病 母;高血圧

【生 活 歴】Alcohol Social drinker;

Smoking (20 本/日, 40 年間)

【輸 血 歴】なし

【入院時現症】

〈GENERAL STATUS & VITAL SIGNS〉BH 165cm,

BW 65kg, BP 112/74, HR 56/min. Cons clear

〈LUNG〉no rales, no wheezing

〈HEART〉no murmur, heart sound regular

〈ABDOMEN〉soft & flat, bowel sound good

〈EXTREMITIES〉no edema

【入院時検査所見】

〈BLOOD CHEMISTRY〉TP 7.0g/dl, Alb 4.1g/dl,

LDH 322IU/l, GOT 84IU/l, GPT 60IU/l, CPK 212IU/l,

T.Bil. 0.8mg/dl, Ca 9.0mg/dl, BUN 18mg/dl, Ce 0.8mg/dl,

Na 140mEq/l, K4.0mEq/l, Cl 108mEq/l, Glu 116mg/dl,

〈CBC〉WBC 9.8x10³/μl, RBC 520x10⁴/ml, Hb 14.6g/dl,

Ht 44.3%, Plt 22.5x10⁴/μl

〈SEROLOGICAL TEST〉CRP 1.8mg/dl, HCV-Ab (-)

〈ABG〉pH 7.430, pCO₂ 34.2mmHg, pO₂ 84.4mmHg,

HCO₃- 24.7mEq/l, BE -0.8, O₂Sat 96.4%

〈Chest X-p〉slight cardiomegaly, no pulmonary congestion
〈ECG〉sinus rhythm, V1 から V6 ST 上昇

【入院後の経過】

ECG および心エコー所見より前壁中隔梗塞を疑い緊急 CAG を施行した。その結果左冠動脈の seg.7 で total occlusion があるものの、右冠動脈から LAD への側副血行路が認められ、右冠動脈には狭窄は認められなかった。AMI と診断し、病変部位に対して再灌流療法(PTCA)を施行した。末梢への血流は良好に再開し、残存狭窄は 20%であった。その後、CCU では再灌流障害も認められず、血圧 120/80、心拍数 80。第3病日での ECG は V1~V5 で Q 波(+), CPK は 230 まで低下(最大 850)。第7病日で CCU から一般病室に移り、2週間のリハビリを良好に経過後、再度の CAG では再狭窄は認められず、退院前のトレッドミル検査では最大心拍数 110、血圧 150/88、ECG 上の ST 変化および胸痛等も認められなかった。以後、外来通院となった。

【考 察】

本症例の経過から AMI 発症数日前より、不安定狭心症を発症していたことが考えられる。このように胸部症状を訴えてきた患者に対しては、その検査所見にかかわらず AMI の可能性を念頭に入れて対処すべきである。

また鑑別診断は、狭心症、急性心膜炎、不整脈、心臓神経症などの心疾患、肺梗塞、自然気胸、急性肺炎、胸膜炎などの呼吸器疾患、その他、解離性大動脈瘤、胆石症、胃十二指腸潰瘍などであり、これらの疾患も考慮に入れた精査を進めていく必要がある。

【退院時処方】

小児用パファリン 1T 1×, フランドル 2T 2×, ヘルベッサー 2T 2×, メバロチン(10mg) 1T 1×