

自然言語処理と アソシアトロンを用いた疾患系分類

5G-3

納富 一宏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) 早稲田大学人間総合研究センター
3) 早稲田大学理工学総合研究センター

1. はじめに

臨床症例文書のデータベース化は、過去の症例を参照し、新たな症例の病態を検討・類推する上で非常に重要である。

我々は 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っている¹⁾。本稿では、自然言語処理とアソシアトロン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。

2. システム構成

2.1 WWW の利用

代表的なインターネットサービスである WWW を利用することは、症例の検索・登録において全国的な地域をカバーするためには大変有利である。また、一般的な WWW ブラウザを用いたデータ検索・閲覧は、その簡便性によりユーザへの操作負担を軽減することが可能である。システム構成を図 1 に示す。

2.2 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、キーワード指定を必要とするため、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、キーワード指定を自動化することを目標とした。このことで、新規症例報告に用いられた文書自体を検索キーとして指定することが可能となった。

2.3 データベース

自然言語文として入力された臨床症例データを元

に、自動インデキシングを通してフルテキストデータベースを構築する。臨床症例データは、A4 で数ページ（通常、1,000～3,000 文字程度）の文書であり、ここからデータベースに登録するための検索用インデックスを自動生成する。

また、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キーワードによる検索、の3つの形式をサポートしている。

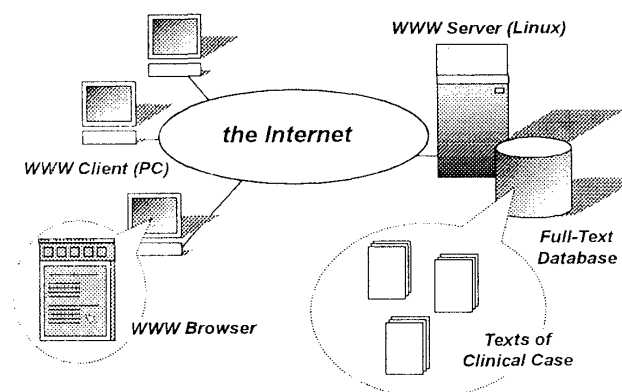


図1. システム構成

3. 自然言語処理

3.1 キーワード抽出

文書データの登録・検索において共通なのが、キーワード抽出処理である。ここでは、一般的な HTML フォーム送受信に利用される CGI (Common Gateway Interface) を用いてサーバ/クライアント間の通信を行う。漢字コードはシフト JIS に統一している。

また、送信された症例文書から、キーワード候

補を選択するために、文字種別に着目した簡易形態素解析を行って自立語候補を抽出する。さらに、

抽出された語集合から、表記文字に関するヒューリスティックスを用いて医学用語候補を選択する。

3.2 データ登録・検索

症例データの登録では、自動抽出された医学用語からなるキーワード集合からインデックス生成を行った後、系分類および実際の登録を行う。

データベースに登録される症例データは、アソシアトロンにより確信度を求め、3つの系(①循環器系、②呼吸器系、③消化器系)に分類した後、それぞれの症例データ領域へ追加する。

検索部では、データベースに存在するインデックス情報と、入力文書から得たキーワード群とを照合し、一致度(%)を求めて、候補を絞り込む。

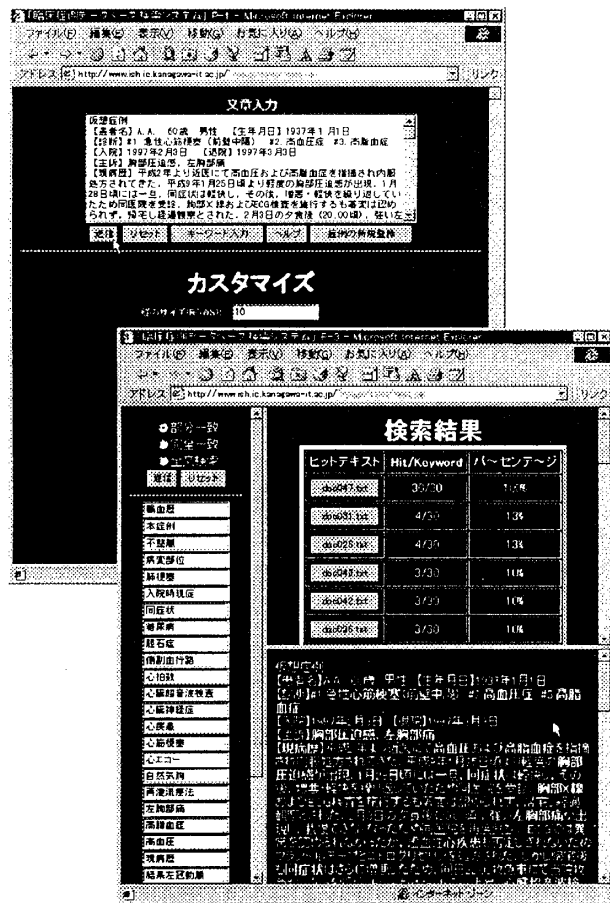


図2. 動作画面例:

症例入力(背面), 検索結果表示(前面), 自動抽出キーワード群(前面・左), 検索結果(前面・右上), 文書表示画面(前面・下)。

4. アソシアトロンによる系分類

本システムは、UNIX ベースの WWW サーバ上に CGI プログラムとして実装した。クライアント側の動作画面例を図2に示す。また、総症例数 35、医師によりあらかじめ定義された系分類用キーワ

ード群 120 個におけるアソシアトロンを用いたパターン想起の様子を図3に、系分類結果を図4にそれぞれ示す。

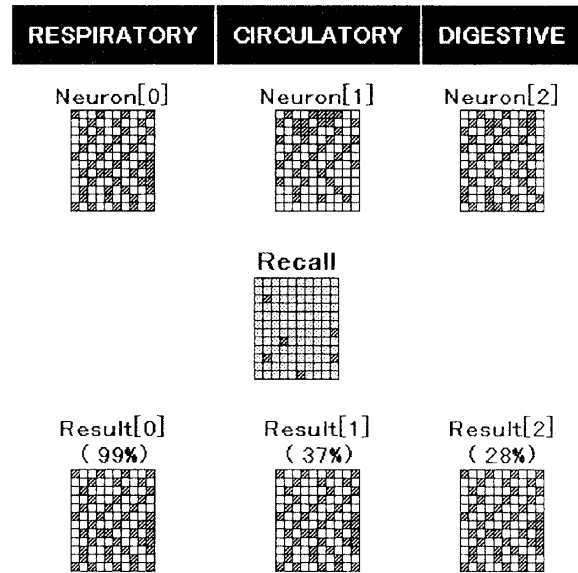


図3. キーワードによる想起例: 銘記パターン(上3つ), キーワードパターン(中), 想起パターンと一致度(下3つ)。

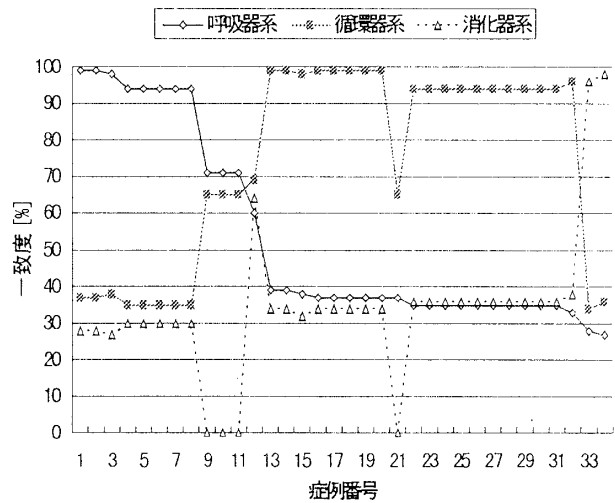


図4.アソシアトロンによる系分類

5. まとめ

自然言語処理とアソシアトロンを用いた疾患系分類について述べた。現在、学習アルゴリズム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).