

外22-32

早稲田大学大学院理工学研究科

博 士 論 文 概 要

論 文 題 目

臨床症例文書を対象とした疾患系分類と
症例データベース構築に関する研究

Research on the Disease System Classification and
Database Construction for Clinical Case Documents

申 請 者

納富

一宏

Kazuhiro

Notomi

2002 年 12 月



医療情報分野における情報共有は、新たな病態の把握や問題点の考察を行なう上で非常に重要である。これには、臨床医学を中心とした各種データの取り扱いが考えられるが、病院をはじめとする各種医療施設で開催される「症例検討会」、「症例報告会」に提出される文書資料を「臨床症例報告書」、または「臨床症例文書」と呼ぶ。

臨床症例は、カルテと同様に患者の個人情報、医師による診断に加えて、既往歴や各種検査・処置・投薬情報などが含まれる定型的な文書形態を有し、通常は日本語で記述される。この症例文書情報の管理を主な目的として、近年、データベース化のニーズが高まっている。これを症例データベースと呼ぶ。

このような臨床症例データを扱うシステムは、単純なインデックスのみで構成される文献データベースや図書データベースといった従来方式の文書管理システムと同列に考えることは困難である。また、利用者層が、基本的に医療に携わる人間、すなわち医師に限定される点が特徴である。文献や図書などにジャンルがあるように、臨床症例にもまた疾患系という分類カテゴリが存在する。これを正しく判断し、以後の類似症例検索時の状況を想定して正しいインデックス用のキーワードを選定し、新規症例をデータベースに追加登録する。当然、こうした手順を踏む際のわずらわしさを極力排除していかなければならない。

現在、インターネットが普及したことで、「検索エンジン」と呼ばれるフリーキーワードによるサイト検索サービスが一般にも広く利用されている。利便性を重視する場合、ネットワークシステムの活用が不可欠であるが、医療情報共有という観点からは、病院などの医療施設の外部に対して臨床症例文書が公開されることは、現状ではまだ多くの問題点を含んでいる。特に、セキュリティの確保が重要な課題となっている。

以上のような背景や問題点を踏まえた上で、本研究の目的は、医療情報分野における情報共有を可能にするための手法の開発、および臨床症例データベースシステムの実現である。また、簡便性、利便性、可搬性、安全性、信頼性の5つのポイントを重視し、実用レベルでのシステム構築を目指す。

本論文は全9章から構成される。以下、各章の概要について述べる。

第1章「緒言」では、本研究の位置付けとして、研究背景および研究目的を明示した。また、論文の全体構成に関して各章の概要について述べた。

第2章「臨床症例データベース」では、本研究で扱う臨床症例報告文書の形式的な定義と、データベース化の必要性を示し、文書検索におけるシステム要件として、症例データベースの基本システム構成を中心に述べた。特に、新規症例の登録、および類似症例の検索の双方について、自然言語処理技術を応用することで、自動インデキシング、および全文検索に対応する形態が重要となる点を示した。また、疾患系による臨床症例の分類については、詳細を第4章、第6章、第7章で示すこととし、ここでは要件項目の概略を示すにとどめた。

第3章「ネットワーク技術と医療情報共有」では、情報共有を実現するために必要不可欠な情報インフラとしてのインターネットの概略に触れると共に、臨床症例データベースのシステム設計に関連する情報提供手段としての Web アプリケーション技術の活用、およびオブジェクト指向とシステム実装に関する従来方式について述べた。また、コンテンツのデータ表現形式としては、XML (eXtensible Markup Language) のサブセットである MML (Medical Markup Language) を従来研究の具体例として取り上げた。

第4章「自然言語処理技術」では、症例文書からの情報取得手段として自然言語処理技術の応用が必要不可欠である点を明らかにし、実用レベルでの医学関連用語の抽出を目的とする基本解析実現のための汎用性の高い形式文節表現モデルとして「J F K 構造」の提案を行なった。さらに、データベース構築の自動化を考える上で必要なキーワード出現頻度による自動インデキシング手法を提案し、具体的なシステム構成、および実装手法について述べた。これにより、フリーキーワードによる全文検索に加え、自然言語（日本語）インタフェースを搭載した検索システムを実現した。また、実現システムによるインターネットを利用した性能評価実験を行ない、実用上問題ない応答時間で動作することを確認することで、検索システムとしての有効性を示した。

第5章「知的文書処理技術」では、記述された情報から、直接的に医療には関係ない（医師としての専門知識が必要とされない）部分での、より一般的な言語表現上の表記誤りやその訂正方式について、日本語文章の校正支援を中心に、手法の提案と実装、および基礎的な評価実験結果を示し、利用可能性の検証を行なった。具体的には、(1)文書校正処理と統語的誤り検出の高速化、(2)専門用語解析における複合語展開と単文化、(3)共起格情報による助詞要素の訂正、(4)連体修飾要素の意味構造解析、の4つの処理方式を取り上げた。本研究で示した知的文書処理方式においては、実装コストを考慮した上で、誤り検出・訂正といった比較的表層レベルの支援について検討したが、限定された適用範囲内では有効であることを確認した。また、今後はより深いレベルでの意味処理が必要となることを示した。

第6章「連想記憶モデル」では、臨床症例文書の分類が検索性の向上に必要であり、特に近年注目されるネットワークシステムの分散・協調について考慮した場合、データベースがあらかじめ疾患系により分類されている方が、システム構成上有利であるという観点から、臨床症例文書の疾患系の判定と分類について、ニューラルネットワークモデルの一つであるアソシアトロン (Associatron) を利用する手法を提案し、併せて分類性能の評価実験を行なった。具体的には、実際に症例検討会で示された症例文書を用い、循環器系、消化器系、呼吸器系の3疾患系に自動分類する手法を示した。また、システムによる分類が、医師による分類に一致する割合を求めることで、本手法が有効であることを明らかにした。

第7章「自己組織化マップ」では、第6章に続くもうひとつのニューラルネットワークモデルである自己組織化マップ(SOM: Self-Organizing Map)を利用した疾患系の判定、および文書クラスタリング方式を提案した。さらに SOM により生成された2次元マップの直観的なデータ把握を可能とする情報視覚化、および検索インタフェースとしての応用例を示し、システムの汎用化と拡張可能性について述べた。

具体的には、第4章で提案した医学関連用語(キーワード)の自動抽出手法を利用し、臨床症例文書の特徴量抽出を行った上で、SOMによる文書クラスタリングを行なう手法を提案した。また、提案手法に基づくシステム構成を示し、症例文書クラスタリングシステムの実現を行なった。さらに、疾患系分類性能の評価実験を行なうと共に、SOMアルゴリズムの適用に際し、ニューラルネットの学習条件として、マップ生成時のパラメータの最適な組み合わせを求めた。パラメータとしては、学習回数、属性数、マップサイズの3点を用いた。第6章と同様、実際の臨床症例報告文書を用い、システムによる3疾患系(循環器系、消化器系、呼吸器系)への自動分類結果が医師による分類に一致する割合を求めることで、本システムの有効性を確認した。

第8章「バイオメトリクス認証」では、不正ユーザの排除を目的として、正規ユーザのパスワード打鍵タイミングパターンを SOM によりあらかじめ分類・登録しておき、システムへのログイン時の打鍵タイミングが登録ユーザによるものか否かをパターンの類似度から判別する個人認証手法を提案した。単一マップによる認証に加え、複数マップによる多数決モデルを用いた認証手法についても検討した。

また、被験者を用いたパスワード盗用時の実証実験から、セキュリティ強化に関して本手法の有効性を評価した。安全性の確保は、プライベート情報を多分に有する臨床症例データベースには必要不可欠な要素である。本研究では、第7章で実現した文書クラスタリング用 SOM エンジンを利用することで、ソフトウェアのみによるバイオメトリクス認証を実現しており、コスト面で非常に有利であることを明らかにした。

第9章「結言」では、第2章から第8章で述べた主要な成果について総括し、本論文のまとめを示した。

以上のように、本研究では、臨床症例文書をターゲットとして、自然言語処理技術およびニューラルネットワークモデルを応用した疾患系分類方式を開発し、医療情報共有を目的とした症例データベースシステムを実現した。さらに Web アプリケーションとしての利便性に加え、コンピュータネットワークのセキュリティ面を考慮することで、最も基本的なシステムへのログイン場面における安全性向上を果たした。これにより、本研究は、実用レベルでの臨床データベース構築手法の確立に貢献した。