

氏 名 石橋 雄一

授与した学位 博士

専攻分野の名称 学 術

学位授与番号 博甲第4441号

学位授与の日付 平成23年 9月30日

学位授与の要件 環境学研究科 資源循環学専攻

(学位規則第5条第1項該当)

学位論文の題目 画像特徴量とテキスト情報による画像検索可能な病情報データベースの構築

論文審査委員 教授 栗原 考次 教授 垂水 共之 教授 梶原 毅

学位論文内容の要旨

類似画像検索の機能を持つ病情報データベースを開発した。データベースには現時点では、乳腺疾患に関する病理組織標本画像や病理診断所見のテキスト情報などを蓄積している。画像検索では、画像情報のみで検索された類似の画像の表示と、検索画像の可能性のある疾患の確率を計算することができる。

画像検索を行うためには、画像を数値化する必要がある。画像を複数の小画像に分割し、ウェーブレット変換することにより数値化を行った。訓練データを疾患の特徴的な部分の小画像とし、ニューラルネットワークによるパターン認識を行うことにより、特徴ベクトルを作成した。検索すべき画像の特徴ベクトルと、データベースに登録されている画像の特徴ベクトルをもとに類似画像検索を行うが、非階層クラスタ分析により分類することで類似画像検索を行う。

病理診断所見には、病理医が標本を観察し診断に至る過程が詳細に記述されている。この診断情報をテキストマイニングにより、キーワードを抽出し辞書を作成することにより、テキストを数値化した。疾患を群として、キーワードを変数とした正準判別分析により、どのキーワードが疾患の判別に係わっているのかを分析した。この結果をもとにベイズの定理を用いて、いくつかのキーワードを与えることにより、疾患ごとの可能性の確率を計算できるようにした。さらに訓練データ画像と訓練データを表現するようなキーワードを結びつけることにより、検索画像に対する疾患ごとの可能性の確率が計算できるようにした。

類似画像検索の精度を向上させるためには、適切な訓練データを選択する必要がある。このため、前述のテキストマイニングで得られた疾患に係わるキーワードを参考にして特徴画像を訓練データとして選択した。これによって得られた訓練データを主成分分析で分析することにより、異なる種類の訓練データで似ているものや、一つの種類の訓練データの中で離れたものなどを検出することができる。これをもとにして、訓練データを整理統合し、不足しているものを追加することにより、パターン認識において誤判別を減らすようにした。

論文審査結果の要旨

本論文では、病理画像情報及び診断書に記述されるテキスト情報に基づく病理組織診断について考察している。

病理組織診断では、癌などの疑いのある生体組織の一部を標本として取り出し、顕微鏡で対象組織を観察し診断や病態の把握及び治療方針の決定などが行われる。一症例あたり数枚から数十枚のプレパラート上の病理組織標本が作製され、それを基に病理医は病理所見やコメントなどを病理診断書に記入する。こうした、病理組織標本や病理診断書は大量に蓄積されていたにも関わらず、その有効利用はほとんどなされていないのが現状である。そこで本研究では、乳腺疾患を対象として病理組織標本画像及び病理診断書における画像情報及びテキスト情報の効率的なデータベース化を行うと共に、診断支援のために画像情報に加えテキスト情報に基づく類似症例の検索が可能なシステムを構築している。

画像情報については、標本画像の特徴的な部分を含む小画像に対して Wavelet 変換により特徴量を 18 変量に数値化し、ニューラルネットワークを用いて 13 種類の疾患に対する特徴ベクトルに基づくパターン認識を行っている。テキスト情報については、病理診断書に記述されている病理所見に対して、単語頻度解析及び係り受け頻度解析によりキーワードと述語の抽出し、キーワードや肯定・否定表現を数値化するための変換辞書を独自に開発している。また、正準判別分析を行い疾患の判別を行うと共にキーワードを与えたときの疾患に対する事後確率をベイズの定理により計算しキーワードと疾患の関連を調べている。これらの研究成果に基づき、画像情報及びテキスト情報をデータベース化すると共に、検索機能を持つ病理情報システムを構築している。画像検索では、検索対象の特徴ベクトルを計算し、k-means クラスタリング手法を用いてデータベースに登録画像の中で類似している画像を表示する。テキスト検索では、キーワードを基に各疾患に対する診断確率を計算することができる。本研究で開発したシステムは、病理組織診断時に既存の診断を容易に参照でき、教育用の利用のみならず、診断の精度の向上を高めることが可能である。

これらの研究成果は、2 編の査読付き論文、2 編の国際会議論文、7 件の口頭発表、1 件の特許出願などにおいて全て第 1 筆者として公表している。これらの研究は、医学統計学及び計算機統計学における理論と応用への貢献は大きい。以上により、本論文は博士（学術）に値すると判断した。