

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年5月30日現在

機関番号：14301

研究種目：若手研究（B）

研究期間：2010～2011

課題番号：22700257

研究課題名（和文） データマイニング手法を用いた類似診療情報検索システムの構築

研究課題名（英文） Development of Retrieval Systems for Similar Medical Record Using Data Mining Methods

## 研究代表者

岡本 和也（OKAMOTO KAZUYA）

京都大学・医学研究科・助教

研究者番号：60565018

## 研究成果の概要（和文）：

本研究では、蓄積された診療情報を用いた類似診療情報検索システムの実現を目指した。(1) 類似症例を発見することで自動DPCコーディングを行う自動DPCコーディングアルゴリズムの開発を行い、従来の結果に比べ高精度なアルゴリズムを開発した。(2) 画像診断レポートの画像タグ情報を利用した類似症例検索システムの開発を行った。(3) 類似の診療情報を紡ぐことで臨床知識獲得のための仮説発見を支援するシステムを構築し、有用性を確認した。

## 研究成果の概要（英文）：

In this study, we developed retrieval systems for similar medical records using accumulated medical information. (1) We developed and evaluated a method of selecting accurate DPC codes by finding similar cases. The study confirmed the feasibility of automatic DPC code selection through machine learning methods. (2) A similar case retrieval exploiting meta-information was realized by giving meta-information of anatomical positions. (3) The authors developed a support system for finding hypotheses about clinical knowledge by narrowing based on similar medical information.

## 交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	600,000	180,000	780,000
2011年度	600,000	180,000	780,000
年度			
年度			
年度			
総計	1,200,000	360,000	1,560,000

研究分野：総合領域

科研費の分科・細目：情報学・図書館情報学・人文社会情報学

キーワード：医療情報、情報検索システム

## 1. 研究開始当初の背景

医療分野では情報技術の適応による診療の効率化や医療の質の向上が期待されている。しかしながら、十分な情報技術の活用が行えていないという認識が一般的であり、2009

年に策定された i-Japan 戦略 2015 においても三大重点分野の1つとして、医療・健康分野が挙げられていた。京都大学医学部附属病院でも電子カルテを用いた診療が行われているが、蓄積された診療情報の利用がほとん

どなされていない。

医師が診療を行う際に、他の似た症例の患者の診療情報を参照することができれば、診療の質を向上させることができると考えられる。そのため、効率的な類似診療情報検索システムは診療業務の効率化のために有用である。そこで、以前に我々は診療文書の特徴を利用した類似診療文書検索システムの提案を行い、医療に特化した効率的な検索システムの実現可能性に関して前向きな結果を得た。但し、その際には診療文書のみを解析の対象として検索を行っていた。診療情報は診療文書の他にも、検査データや処置データといった様々な情報を含むため、これらを解析の対象として利用できれば、検索精度の向上が見込まれる。

## 2. 研究の目的

医師による診療の質の向上のため、類似診療情報検索システムを構築することを目的とする。診療情報は診療文書、検査データ、処置データ等からなる。診療文書の解析を行うツールとして我々が以前に提案した診療文書の構造を利用した類似診療文書検索技術を利用する。また、検査データや処置データの解析を行うため、サポートベクターマシンを始めとする各種データマイニング技術を応用する。そして、これらの解析結果を組み合わせることによって、効率的な類似診療情報検索を行うための手法の確立を目指す。具体的な取り組み項目を以下に挙げる。

## 3. 研究の方法

診療文書の構造を利用した類似診療文書検索技術に組み合わせる検査データ・処置データからの類似性算出手法の開発を行う。検査データに関しては、前述したように単純に数値の近さでは類似性が測れないこともある。そこで、まず正常値・異常値といった情報を基に類似度を測定する。その上で、重要な検査データに絞り、正常値における、あるいは、異常値における数値の近さに意味があるか分析し、類似度の測定に組み込むことを試みる。また、処置データに関しては、各処置データの関連性をデータマイニングツールで分析し、分析結果を基に類似性を算出する。続いて、診療文書の構造を利用した類似診療文書検索技術に検査データ・処置データからの類似性算出手法を組み合わせ類似診療情報検索システムを構築する。

## 4. 研究成果

(1) 類似診療情報検索システムの中心のアルゴリズムになりうる自動DPCコーディングアルゴリズムの開発を行った。DPCコードは主傷病に加えて、副傷病や処置等を表すコードである。2010年現在、1334の病院等にお

いてDPCコードが診療報酬の決定に用いられている。DPC対象病院において各症例の診療報酬は1つのDPCコードから決定されるため、各症例を1つのDPCコードに割り当てなければならない。しかし、患者は多様な容態を示すため、1つのDPCコードに割り当てることは困難である。そこで、機械学習による自動DPCコーディングの可能性の検討が行われている。これまでに、Suzukiらは退院時サマリを解析対象としてベクトル空間法を用いた自動DPCコーディングを試みた。しかし、ベクトル空間法以外にも一般的な機械学習手法は複数存在し、自動DPCコーディングの精度向上のために他の機械学習手法を網羅的に検討する必要がある。そのため、我々は同じく退院時サマリを解析対象とし、一般的な機械学習手法であるナイーブ・ベイズ法、SVM、概念ベース法、さらに、既存研究で用いられたベクトル空間法と異なるベクトル空間法（平均ベクトル空間法）を用いて自動DPCコーディングを行い、それらの精度の比較を行った。また、我々は自動DPCコーディングにおいて各々の機械学習手法に特徴があることに注目し、各機械学習手法が分類の根拠としているそれぞれの分類スコアから分類結果に対する各機械学習手法の確信度を推定し、確信度の最も高い結果を選択する方法、すなわち、確信度の最大化に基づいて複数の機械学習手法を統合する手法を試みた。そして、それらの精度の比較を2008年度に京都大学医学部附属病院において作成された退院時サマリを用いて行った。その結果、Suzukiらが用いたベクトル空間法では64.8%の症例が正解DPCコードに割り当てられたのに対し、SVMでは72.2%の症例が正解DPCコードに割り当てられた。また、統合手法では76.1%の症例が正解DPCコードに割り当てられた。結果として、大幅な精度の向上を達成した。また、入院中に行われる処置の特徴として、手術などの特徴的な処置は入院初期に行われる、また、段階的に処置が行われるという特徴があると考えられる。そのため、入院初期に行われた処置から自動DPCコーディングを行うことで、効率的なコーディングが可能であると思われる。さらに、入院初期までのデータを分類する時に、入院後期でしか行われない処置を含む教師データで学習したモデルを用いることは適切ではないと思われる。そこで、本研究では、入院期間に応じた自動DPCコーディングの精度を評価する。結果として、さらに精度を向上させることに成功した。

(2) 画像診断レポートは、基本的に画像のもつ情報と記述されたテキスト情報の組み合わせでひとつの情報であるといえる。しかし、これらの情報を結びつけてレポート情報を

再利用する方策はまだ採られていない。

我々は、レポート情報のより有効な再利用と類似症例検索への適応を目指し、画像情報のメタ情報およびアノテーション情報を利用する方法論を構築した。その上で、これを類似症例検索に利用するシステムの構築を試みた。

具体的には、シェーマに対してどのようなシェーマなのかというメタ情報、および紙媒体のレポートで行われていたある文章が指し示す位置（座標）情報が入力可能で、かつそれらの情報も加味した類似症例検索が可能なシステムを構築し、実際のレポート情報を入力して有用性を検証した。

(3) 診療情報を用いた、臨床知識獲得のための仮説発見を支援するシステムを構築し、構築システムの評価を行う。構築システムは検索対象となる診療情報全体、および、検索により絞り込まれた診療情報を俯瞰的に提示する。ユーザは提示された情報を用いて診療情報を絞り込むことができる。これら「システムによる俯瞰的な診療情報の提示」と「提示された情報を用いたユーザによる診療情報の絞り込み」によりインタラクティブな検索が実現される。インタラクティブな検索によって、診療情報の絞り込みに用いるクエリの生成を支援し、ユーザの求める仮説の発見を支援する。

本研究では、退院時サマリ、血液検査情報、病名情報、および、処置・手術・処方等の医事情報といった診療情報をユーザに俯瞰的に提示することで、臨床的な知識に繋がる仮説の発見を支援するシステムの構築を行った。まず、症例集合が持つ診療情報を集計し、集計結果をスニペット・グラフ・タグクラウドによって俯瞰的に提示することで、ユーザの曖昧な臨床知識に対する仮説を明確化する支援を行った。次に、ユーザが提示情報をもとに検索条件を追加できるようにし、インタラクティブな検索によって症例集合を自身の要求を的確に満たす症例集合に絞り込むことができる仕組みを構築した。

構築システムを医師に利用してもらうことで構築システムの情報提示に対する有用性の検証と臨床知識に繋がる仮説発見支援の可能性の検証を行った。医師の評価より、ユーザが明確な要求を持っていない場合に、網羅的な情報提示が新たなクエリを決める助けとなることがわかった。また、診療情報を俯瞰的に見ることができ、直感的にも分かりやすく面白いと、診療情報を見ること自体に興味を持ちやすくなるという評価が得られた。これらの評価により、構築システムは臨床知識の獲得に繋がる仮説発見支援に有用であると考えられる。

## 5. 主な発表論文等

〔雑誌論文〕（計7件）

- ① Minoru Yamada, Tomoki Aoyama, Shuhei Mori, Shu Nishiguchi, Kazuya Okamoto, Tatsuaki Ito, Shinyo Muto, Tatsuya Ishihara, Hiroyuki Yoshitomi and Hiromu Ito, Objective assessment of abnormal gait in patients with rheumatoid arthritis using a smartphone, *Rheumatology International*, 査読有、印刷中、2012、DOI: 10.1007/s00296-011-2283-2、
- ② 岡本和也、田中寛子、竹村匡正、糸直人、黒田知宏、吉原博幸、臨床知識獲得のための診療情報を用いた仮説発見支援システムの構築、医療情報学、査読有、印刷中、2012、
- ③ Tadamasa Takemura, Kenji Araki, Kenji Arita, Toshiaki Suzuki, Kazuya Okamoto, Naoto Kume, Tomohiro Kuroda, Hiroyuki Yoshihara, Development of Fundamental Infrastructure for Nationwide EHR in Japan, *Journal of Medical Systems*, 査読有、印刷中、2012、DOI: 10.1007/s10916-011-9688-z、
- ④ 江上廣一、廣瀬昌博、竹村匡正、岡本和也、津田佳彦、大濱京子、本田順一、島弘志、今中雄一、吉原博幸、インシデントレポート・医事管理データによる転倒・転落に起因する追加的医療費算出の試み、*日本医療・病院管理学会誌*、査読有、Vol. 48、2011、pp. 33-44、
- ⑤ Minoru Yamada, Tomoki Aoyama, Kazuya Okamoto, Koutatsu Nagai, Buichi Tanaka, and Tadamasa Takemura, Using a Smartphone while walking: a measure of dual-tasking ability as a falls risk assessment tool, *Age Aging*, 査読有、Vol. 40、2011、pp. 516-519、DOI: 10.1093/ageing/afr039、
- ⑥ 竹村匡正、田中陽子、岡本和也、糸直人、黒田知宏、廣瀬昌博、吉原博幸、ロールプレイを用いた病院モデルに基づく教育用病院経営シミュレーションゲームの構築、医療情報学、査読有、Vol. 30、2011、pp. 37-48、
- ⑦ 岡本和也、内山俊郎、竹村匡正、足立貴行、糸直人、黒田知宏、内山匡、吉原博幸、複数の機械学習手法を用いた退院時サマリからの自動DPCコーディング、*生体医工学*、査読有、Vol. 49、No. 1、2011、pp. 40-47、

〔学会発表〕（計8件）

- ① 岡本和也、竹村匡正、村田節子、糸直人、黒田知宏、吉原博幸、記入型ライフログ

- 蓄積サービスを用いたライフログの診療への活用の取り組みと課題、医療情報学連合大会、2011/11/22、鹿児島、
- ② 岡本和也、内山俊郎、竹村匡正、足立貴行、糸直人、黒田知宏、内山匡、吉原博幸、入院期間に応じた処置情報からの自動 DPC コーディング、医療情報学連合大会、2011/11/22、鹿児島、
- ③ 岡本和也、廣瀬昌博、竹村匡正、江上廣一、中林愛恵、西村信弘、津田佳彦、本田順一、今中雄一、インシデントレポートの精緻化および自動化の試み、日本医療・病院管理学会、2011/8/21、東京、
- ④ Kazuya Okamoto, Tadamasa Takemura, Naoto Kume, Tomohiro Kuroda, Naoki Oboshi, Kenta Hori, Hirose Masahiro, Nobuaki Ito, and Hiroyuki Yoshihara, A Reporting System and a Similar Case Retrieval System Exploiting Meta-information of Image、CARS、2011/6/22、Berlin/Germany、
- ⑤ 岡本和也、田中寛子、竹村匡正、糸直人、黒田知宏、吉原博幸、診療情報を対象とした臨床知識獲得支援システムの構築、医療情報学会春期学術大会、2011/6/18、千葉、
- ⑥ Shuichi Miyazaki, Kazuya Okamoto、Improving the Competitive Ratios of the Seat Reservation Problem、6th IFIP TC 1/WC 2.2 International Conference, TCS 2010、2010/9/23、Brisbane/Australia、
- ⑦ 岡本和也、内山俊郎、竹村匡正、足立貴行、糸直人、黒田知宏、内山匡、吉原博幸、複数の機械学習手法を用いた退院時サマリからの自動 DPC コーディング、生体医工学シンポジウム 2010、2010/9/11、北海道札幌市、
- ⑧ 内山俊郎、岡本和也、竹村匡正、足立貴行、糸直人、黒田知宏、内山匡、吉原博幸、医療分野に特化した概念ベースによる DPC コーディング、電子情報通信学会パターン認識・メディア理解研究会、2010/5/14、愛知県春日井市、

[その他]

第 31 回医療情報学連合大会若手奨励賞受賞

## 6. 研究組織

### (1) 研究代表者

岡本 和也 (OKAMOTO KAZUYA)

京都大学・医学研究科・助教

研究者番号：60565018