

科学研究費助成事業 研究成果報告書

平成 30 年 6 月 9 日現在

機関番号：14301

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00466

研究課題名(和文) インタラクティブ検索を利用した診療情報からの仮説生成支援環境の構築

研究課題名(英文) Construction of an environment to support producing hypotheses from clinical information by using interactive retrieval

研究代表者

岡本 和也 (Okamoto, Kazuya)

京都大学・医学研究科・准教授

研究者番号：60565018

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：本研究では、様々な種類の診療情報を網羅的に扱い、研究者に各種診療情報を俯瞰的に見せることで、研究者が気づきという仮説を発見できる環境の構築を行った。そのために(i)診療情報における重要項目の算出、(ii)診療文書からの情報抽出、(iii)診療情報の俯瞰的提示手法の開発、(iv)仮説生成支援環境の構築、という4つのテーマを設定した。そして、それぞれのテーマについて課題を設定し、解決策を提示した。

研究成果の概要(英文)：In this research, we treated various types of clinical information comprehensively, and showed researchers various kinds of clinical information in a bird's-eye view, thereby constructing an environment in which researchers can discover hypotheses of awareness. In order to construct the mentioned environment, we set four themes: (i) calculation of important items in clinical information, (ii) information extraction from clinical documents, (iii) development of bird's-eye presentation method of medical information, (iv) creation of hypothesis generation support environment. Then, we set issues on each theme and presented solutions.

研究分野：医療情報学

キーワード：データマイニング 機械学習 診療情報

1. 研究開始当初の背景

病院への電子カルテの導入が進み、診療情報が電子的に蓄積されてきている。診療情報の電子化の利点として、診療情報の共有、医療安全への応用などがあるが、同様に知識処理による診療情報の活用が期待されている。例えば、診療記録の記載から有害事象を抽出する研究は国内外で盛んに行われている。また、糖尿病などの特定の疾患に注目し、関連する診療情報を抽出した上で解析をするという試みも広く行われている。このようなニーズに対して、研究代表者は若手研究(B)「Web サービスを利用した診療情報検索システムの構築」(H21~H23)において、各種診療情報を提供する Web サービスを構築し、診療情報の利用を促進してきた。このように蓄積された診療情報の活用が進んでいるが、有害事象の抽出のように、利用対象となる診療情報はごく限られた種類のものであるか、糖尿病に関する解析のように、解析の目的が明確になっており、関連する診療情報のみを解析するというパターンの研究がほとんどである。

2. 研究の目的

医療の世界において、研究者は臨床研究を開始するまでに仮説の生成に非常に労力・時間をかけている。そこで、本研究では、電子カルテの導入に伴い活用することが強く求められている診療情報を利用し、研究者による仮説の生成を支援するための環境を構築する。仮説生成を支援するための技術としてインタラクティブ検索技術を用いることで、検索要求が明確になっていない研究者にも仮説に繋がる気づきを与えることができる。また、インタラクティブ検索に用いるための効果的な診療情報の俯瞰的提示手法の開発を行うことも本研究の目的となる。

3. 研究の方法

本研究では、様々な種類の診療情報を網羅的に扱い、研究者に各種診療情報を俯瞰的に見せることで、研究者が気づきという仮説を発見できる環境の構築を行う。また、本研究において基礎要素となる研究課題は以下の3つのものである。

(i) 診療情報における重要項目の算出

各種診療情報を網羅的に扱う際に、診療情報ごとの重要度の差異、あるいは、同種の診療情報内にも重要度の差異がある。例えば、血液検査結果などは入院直前、あるいは、入院直後の結果が重要であると考えられ、タイミング情報の利用が有用である。また、検査結果の中には投与薬剤と密接に関連しているものも多い。そのため、これらの重要度を算出するアルゴリズムを開発する。

(ii) 診療文書からの情報抽出

診療情報の中には診療録のように自然言

語で記載された診療文書も多い。そのため、自然言語記載の診療文書から患者の病態を特徴付ける重要事項を抽出する。重要項目の抽出は自然言語処理技術において重要視される「あり」「なし」表現の抽出や単語の重要度だけでなく、(i) 同様に診療情報特有のタイミング情報なども利用する。

(iii) 診療情報の俯瞰的提示手法の開発

各種診療情報を研究者に俯瞰的に提示する手法の開発を行う。例えば数値情報であればグラフを利用することで全体を俯瞰的に提示することができる。同様に、診療情報の中には傷病名などのカテゴリ情報や診療文書などのテキスト情報など様々な形式の情報が含まれている。そのため、数値情報・カテゴリ情報・テキスト情報の俯瞰的提示手法の開発を行い、各種の診療情報に適する俯瞰的提示手法を開発する。

そして、上記要素(i), (ii), (iii)を活用し、研究者が利用できる(iv) 仮説生成支援環境の構築を行う。この仮説生成支援環境を構築するにあたり、本申請研究の方法の鍵となるインタラクティブ検索を利用する。そして、インタラクティブ検索を利用した仮説生成支援環境を用いることで、実際に仮説の生成が可能となるか評価を行う。

4. 研究成果

(i) 診療情報における重要項目の算出

本テーマにおいては、月経前症候群の管理におけるスマートフォンアプリを用いた症状記録システムの開発と臨床使用を行い、各種情報の重要性を確認した。月経前症候群は心身の不快な症状が月経前の黄体期に繰り返し出現する病態であるが、その診断は患者自身による前方視的記録に基づいて月経随伴症状の出現時期と反復性を確認することに依る。したがって月経前症候群を正しく診断し適切に治療するには患者が症状記録を実践することと症状の変調を医師 患者間で確認することが重要である。これを簡便化、効率化する目的で、患者が月経随伴症状やその他の健康情報をスマートフォンで簡便に入力できるアプリと、その内容を病院内のサーバに保管し医師、患者ともに閲覧できるシステムを開発し、臨床使用を開始した。

また、本テーマにおいて、UGT1A1の遺伝子型と肺癌の罹患との関係を扱い、UGT1A1の遺伝子型という診療情報の重要性を調査した。本研究では、単一施設で2004~2014年に化学療法を受けた5285例のうちUGT1A1の遺伝子型が測定された765例を研究対象とした。肺癌とその他の癌の群に分け、UGT1A1*6及び*28の頻度の差を検証した。劣性・優性・相加のそれぞれのモデルで、ロジスティック回帰を行い検証した。そして、UGT1A1*28*28の頻度(UGT1A1*28劣性モデル)は、肺癌においてその他の癌と比較

して高いことがわかった。なお、その他の遺伝子多型では、肺癌とその他の癌とで差を認めなかった。

(ii) 診療文書からの情報抽出

本テーマにおいては、インシデントレポートを学習データとした電子カルテからのインシデント検出を行った。インシデントの報告に関する問題として、インシデントの報告漏れが挙げられる。インシデントの報告がなければインシデントへの対応を取ることも困難となるため、報告が漏れているインシデントを検出できることは有用である。そこで、本研究では機械学習手法を用いて報告が漏れているインシデントを電子カルテから検出することを目的とした。方法としては、まず、機械学習手法でモデルを作成するために教師データを用意した。教師データは京都大学医学部附属病院において2016年1月1日から2016年6月30日に記載された経過記録のうち、「注射薬」のカテゴリに分類されたインシデントレポートの患者IDと発生日が一致する8,744件の経過記録を用いた。この経過記録のうち、注射薬に関するインシデントの内容を含む279件の経過記録を正例とし、含まない8,465件を負例とした。これらの経過記録を形態素解析器 MeCab で形態素に分割し、「名詞」、「動詞」、「形容詞」、「副詞」、「助動詞」、さらに、単語の組み合わせをベクトル空間の軸とする単語ベクトルに変換して教師データとした。次に、用意した教師データを用いて機械学習手法 SVM (サポートベクターマシン) でモデルを作成した。その際、データの正規化や SVM で用いる各種のカーネルに関する検討も行った。最後に、作成したモデルを用いて2018年1月8日から2018年1月14日までの1週間分の経過記録からインシデント候補の抽出を行った。データの正規化と SVM で用いる各種のカーネルの評価では、教師データを用いて10分割交差検定で評価を行った。結果として、正規化を行った方がより良い結果を得られ、カーネルの中では線形カーネルが最も良い結果を示すことがわかった。この「正規化あり」・「線形カーネル」という設定において再現率が40.1%、精度が88.2%という結果となった。前述の設定を用いることで、対象の経過記録からインシデント候補91件が抽出された。さらに、抽出されたインシデント候補を医療安全管理部のスタッフに確認してもらい、23件がインシデントに関する経過記録であり、その内13件が未報告のインシデントに関する経過記録であった。

また、本テーマにおいて、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 コーディングを行った。DPC 対象病院では、診療報酬請求のために、対象となる各入院患者に対して主傷病に加えて副傷病や処置等を表すコードである DPC コードを1つ付与する必要がある。しかし、多様な容態を示す

患者を1つのコードに割り当てることは困難であることから、診療文書を解析対象とした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。そこで、各種医療情報を用いた自動 DPC コーディングを試み、開発した自動 DPC コーディング手法を利用した自動 DPC コーディング支援システムのプロトタイプを作成した。入院症例のオーダ情報に診療科情報と正解となる DPC コードを付与し、評価実験を行った。評価実験では入院日ごとの精度を測るため、オーダ情報を入院日から x 日目のものに絞って実験を行い、 x 日目の評価とする。 x は入院当日を表す1から入院日から20日目を表す20まで動かす。実験は10分割交差を用いて行った。評価項目として、SVM が第1候補とする DPC コードの全14桁が正解となる DPC コードと一致する割合(全14桁一致精度)、SVM が候補とする5つの DPC コードの全14桁が正解となる DPC コードと一致する割合(全14桁5位以内精度)、SVM が第1候補とする DPC コードの主傷病名を表す上位6桁が正解となる DPC コードと一致する割合(上位6桁一致精度)、SVM が候補とする5つの DPC コードの上位6桁が正解となる DPC コードと一致する割合(上位6桁5位以内精度)を用いた。結果、全14桁一致精度が最も高くなったのは $x=18$ の時であり、78.6%となった。全14桁5位以内精度が最も高くなったのは $x=20$ の時であり、93.3%となった。上位6桁一致精度が最も高くなったのは $x=19$ の時であり、85.1%となった。上位6桁5位以内精度が最も高くなったのは $x=16$ の時であり、96.3%となった。

(iii) 診療情報の俯瞰的提示手法の開発

本テーマにおいては、処方データを解析者が直観的に投薬情報として理解できるように、投薬情報として再構築する手法を構築した。再構築された投薬情報があれば、次の処方方を効率的に判断できると考えられる。しかし、パターン抽出の際に、電子カルテには処方箋の情報しか記録されていないため、本来医師が意図していた処方内容である投薬歴をそのまま把握することは困難である。本研究では、ルールベースで処方箋データから構築した投薬歴に対して、医師の多様性という概念を考慮した投薬パターンマイニング手法を適用し、処方対象患者に対する次の処方薬を分析することで、プライマリケアにおける薬物療法に関する臨床判断支援を行うことを目的とした。京都大学医学部附属病院の電子カルテデータを用いた評価実験の結果、投薬歴に関して、提案手法は既存手法よりも適切な投薬歴が構築できることを示した。また、投薬パターンマイニングに関して、提案手法によって抽出したパターンは、特定の条件下において、既存手法によって抽出したパターンよりも検査値の改善が見込めるパターンを抽出できることを示した。

(iv) 仮説生成支援環境の構築

本テーマにおいては、京都大学医学部附属病院において、医科レセプトデータ・DPCレセプトデータ・調剤レセプトデータを用いた胃癌症例の実態把握を行うため、胃癌症例を対象としたレセプト情報解析のためのデータマートの構築を行った。厚生労働省より提供されるレセプトデータはいくつかの問題点があり、分析に適した形にはなっていない。具体的には、データが各レセプト種別・各人・各月のレセプトを単位として蓄積されているために各人のレセプトデータが分割され、各人を単位としたデータではない、レセプトと各人の紐付けは被保険者証の記号や番号等を基本とした識別子を用いてなされているが、これらの識別子は転居や転職等により変化する、また、レセプトデータの一部に省略があり同一レセプト内の他のデータを参照する必要があるものがあるという問題点が存在する。本研究ではこれらの問題点に対して、各人を単位としたデータに変換し、同一被保険者と推測される識別子の名寄せを行い、省略されているレセプトデータの補完を行うことで、レセプト情報解析のためのデータマートを構築した。

5. 主な発表論文等

〔雑誌論文〕(計5件)

江川美保、岡本和也、西村史朋、森野佐芳梨、糸直人、青山朋樹、小西郁生、月経前症候群の管理におけるスマートフォンアプリを用いた症状記録システムの開発と臨床使用、女性心身医学、査読有、21(1)、2016、105-113

Yoshitaka Nishikawa, Masashi Kanai, Maiko Narahara, Akiko Tamon, JB Brown, Kei Taneishi, Masahiko Nakatsui, Kazuya Okamoto, Yu Uneno, Daisuke Yamaguchi, Teruko Tamono, Yukiko Mori, Shigemi Matsumoto, Yasushi Okuno, Manabu Muto, Association between UGT1A1*28*28 genotype and lung cancer in the Japanese population, International Journal of Clinical Oncology, 査読有, 22(2), 2017, 269-273
doi:10.1007/s10147-016-1061-2

Purnomo Husnul Khotimah, Yuichi Sugiyama, Masatoshi Yoshikawa, Akihiro Hamasaki, Osamu Sugiyama, Kazuya Okamoto, Tomohiro Kuroda, Medication Episode Construction Framework for Retrospective Database Analyses of Patients with Chronic Diseases, IEEE Journal of Biomedical and Health Informatics, 査読有, 2017, to appear
doi: 10.1109/JBHI.2017.2786741

Yuji Morita, Masatoshi Yoshikawa, Noboru Kada, Akihiro Hamasaki, Osamu Sugiyama, Kazuya Okamoto, Tomohiro Kuroda, Medication Pattern Mining Considering Unbiased Frequent Use by Doctors, European Journal for Biomedical Informatics, 査読有, 14(1), 2018, 26-33

Kazuya Okamoto, Toshio Uchiyama, Tadamasa Takemura, Naoto Kume, Tomohiro Kuroda and Hiroyuki Yoshihara, Automatic Selection of Diagnosis Procedure Combination Codes Based on Partial Treatment Data Relative to the Number of Hospitalization Days, European Journal for Biomedical Informatics, 査読有, 14(1), 2018, 34-40

〔学会発表〕(計15件)

松本繁巳、八上全弘、岡本和也、金井雅史、富樫かおり、黒田知宏、武藤学、日常がん診療におけるLMS (Lesion Management Solution, Median technologies, France)を用いた治療効果判定システムの構築、第13回日本臨床腫瘍学会学術集会、2015、P1-4-5

森由希子、岡本和也、加藤源太、杉野剛史、首藤真義、黒田知宏、武藤学、レセプト情報を用いた本邦における胃がん治療の実態調査、第13回日本臨床腫瘍学会学術集会、2015、P2-8-18

采野優、種石慶、金井雅史、多門啓子、岡本和也、西川佳孝、Brown, J.B.、松本繁巳、奥野恭史、武藤学、時系列リアルワールドデータ解析による終末期予後予測モデルの確立、第13回日本臨床腫瘍学会学術集会、2015、O3-5-4

中井隆史、竹村匡正、櫻井理紗、藤田健一郎、岡本和也、黒田知宏、線形サポートベクターマシンを使用した前日の診療行為からの診療行為予測の研究、第35回医療情報学連合大会、2015、268-269

藤田健一郎、竹村匡正、中井隆史、岡本和也、黒田知宏、電子カルテにおけるcontext awarenessの実現に向けた、業務端末以外からの電子カルテ利用状況の分析、第35回医療情報学連合大会、2015、494-497

吉原和音、竹村匡正、藤田健一郎、疋田智子、梶村郁子、岡本和也、黒田知宏、電子カルテシステムにおけるG U Iの現状分析およびユーザビリティの検討、第35回医療情報学連合大会、2015、506-507

岡本和也、森由希子、加藤源太、黒田知宏、武藤学、胃がん患者の治療実態調査のためのレセプトデータ再構築、第35回医療情報学連合大会、2015、506-507

報学連合大会、2015、606-608

山田ひとみ、竹村匡正、朝井隆裕、岡本和也、黒田知宏、奈良崎大士、桑田成規、医療訴訟における説明義務違反と診療録記載に関する考察・最近の判例における事実認定の傾向から、第 35 回医療情報学連合大会、2015、844-845

岡本和也、森由希子、加藤源太、黒田知宏、武藤学、胃癌症例を対象としたレセプト情報解析のためのデータマートの構築、第 53 回日本癌治療学会学術集会、2015、P65-10

武藤学、金井雅史、森由希子、松本繁巳、佐久間朋寛、小柳智義、岡本和也、加藤源太、黒田知宏、森田智視、小杉眞司、多門啓子、種石慶、奥野恭史、高折晃史、京大病院におけるバイオバンク、クリニカルシーケンス、ビッグデータ解析の取組、第 53 回日本癌治療学会学術集会、2015、P50-2

Purnomo Husnul Khotimah, Yuichi Sugiyama, Masatoshi Yoshikawa, Akihiro Hamasaki, Kazuya Okamoto, Tomohiro Kuroda, Revealing Oral Medication Patterns from Reconstructed Long-Term Medication History of Type 2 Diabetes, Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2016, 5599-5603

Purnomo Husnul Khotimah, Masatoshi Yoshikawa, Akihiro Hamasaki, Osamu Sugiyama, Kazuya Okamoto, Tomohiro Kuroda, Comparing Frequent Patterns: A Study Case of Apriori and Singleton Implementations in a Diabetes Type 2 Data Set, Proceedings of International Conference on Computer, Control, Informatics, and its Applications (IC3INA2016), 2016, 163-168

西川佳孝、金井雅史、奈良原舞子、種石慶、岡本和也、采野優、森由希子、松本繁巳、奥野恭史、武藤学、UGT1A1 遺伝子多型と肺癌のリスクに関する研究、医学生・研修医の日本内科学会ことはじめ 2016 東京 プログラム・抄録集、2016、150

岡本和也、五家花梨、廣瀬昌博、中林愛恵、黒田知宏、インシデントレポートを学習データとした電子カルテからのインシデント検出の試み、第 55 回日本医療・病院管理学会 学術総会プログラム、2017、1-B2-03

五家花梨、岡本和也、廣瀬昌博、平木秀輔、山本豪志朗、杉山治、南部雅幸、黒田知宏、データマイニング技術を用いた電子カル

テからのインシデント検出の試み、第 37 回医療情報学連合大会、2017、814-817

〔図書〕(計 0 件)

〔産業財産権〕

○出願状況(計 0 件)

○取得状況(計 0 件)

〔その他〕
該当なし

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・准教授
研究者番号：60565018

(2)研究分担者

黒田 知宏 (KURODA, Tomohiro)
京都大学・医学部附属病院・教授
研究者番号：10304156

田村 寛 (TAMURA, Hiroshi)
京都大学・国際高等教育院・特定教授
研究者番号：40418760

加藤 源太 (KATO, Genta)
京都大学・医学研究科・准教授
研究者番号：20571277

浦西 友樹 (URANISHI, Yuki)
大阪大学・基礎工学研究科・准教授
研究者番号：00533738

(3)連携研究者

該当なし

(4)研究協力者

澤野 良輔 (SAWANO, Ryosuke)
京都大学・大学院情報学研究科・修士学生

Purnomo Husnul Khotimah
京都大学・大学院情報学研究科・博士学生

森田 祐司 (MORITA, Yuji)
京都大学・大学院情報学研究科・修士学生

五家 花梨 (GOKA, Karin)
京都大学・大学院情報学研究科・修士学生