

厚生労働科学研究費補助金

政策科学総合研究事業

(臨床研究等ICT基盤構築研究事業)

電子カルテ情報をセマンティクス(意味・内容)の標準化により分析可能なデータに変換するための研究

平成28年度 総括研究報告書

研究代表者 宮本 恵宏

平成29(2017)年 3月

目 次

I . 総括研究報告

電子カルテ情報をセマンティクス（意味・内容）の標準化により分析可能なデータに変換する ための研究	-----	1
宮本 恵宏		

II . 分担研究報告

1 . 日本循環器学会データ出力標準フォーマットガイドラインに基づく心臓超音波レポート および心臓カテテル検査レポートデータベースの作成	-----	4
中村 文明・西村 邦宏・斉藤 能彦・安田 聡・宮本 恵宏 （資料）添付資料1 心臓カテテル検査レポート設計書 添付資料2 心臓超音波検査レポート設計書		
2 . データ転送プログラムによるデータ収集に関する研究	-----	56
中山 雅晴・興梠 貴英・的場 哲哉		
3 . 臨床効果データベース事業の学会支援に関する研究	-----	58
小室 一成		
4 . 電子カルテ情報のSS-MIX2連携に関する研究	-----	60
宇宿 功市郎・辻田 賢一		
5 . 自然言語解析による診断名判断システムの開発に関する研究	-----	61
宍戸 稔聡・平松 治彦・上村 幸司・竹村 匡正		
III . 研究成果の刊行に関する一覧表	-----	63

平成28年度厚生労働科学研究費補助金
(政策科学総合研究事業(臨床研究等ICT基盤構築研究事業))
総括研究報告書

「電子カルテ情報をセマンティクス(意味・内容)の標準化により分析可能な
データに変換するための研究」

研究代表者	宮本 恵宏	国立循環器病研究センター循環器病統合センター・センター長
分担研究者	竹村 匡正	兵庫県立大学大学院応用情報科学研究科教授
	中村 文明	国立循環器病研究センター循環器病統合情報センター室長
	興梠 貴英	自治医科大学准教授
	中山 雅晴	東北大学病院メディカルITセンター教授
	的場 哲哉	九州大学病院講師
	小室 一成	東京大学大学院医学系研究科教授
	斎藤 能彦	奈良県立医科大学教授
	安田 聡	国立循環器病研究センター副院長・心臓血管内科部門長
	穴戸 稔聡	国立循環器病研究センター研究推進支援部部長
	西村 邦宏	国立循環器病研究センター循環器病統合情報センター室長
	平松 治彦	国立循環器病研究センター情報統括部部長
	上村 幸司	国立循環器病研究センター研究推進支援部研究情報基盤管理室室長
	辻田 賢一	熊本大学大学院生命科学研究部教授
	宇宿 功市郎	熊本大学医学部附属病院教授

研究要旨 日本循環器学会の事業である「臨床効果データベース」は、医療コストがかかる疾患・治療(心筋梗塞・狭心症とその病態に対するステント治療、重症心不全とそれに対する再同期療法(CRT))と循環器領域で特にその重要性が指摘されている疾患(急性心不全など)を抽出し、医療の質とその妥当性を検証するため時間軸を念頭においたデータベースである。本研究ではこのデータベースを用いて自然言語解析による診断判定システムの構築をおこなう。

A. 研究目的

高齢化社会の中にある我が国をはじめとする先進諸国において、循環器疾患が急増している。循環器疾患は再発を繰り返し徐々に進行していくという臨床経過をたどることが多い。例えば、虚血性心疾患では再発・入院を繰り返して終末像として心不全を呈することがしばしばある。そのため循環器疾患においては、Major Adverse Cardiac Event(MACE)とよばれる主要有害心血管イベントを発生させないための再発予防が大事である。循環器疾患の新規治療法の開発目標として、MACEの発生減少を目標としたものを開発することも考えられる

が、MACEを判断するためには担当した臨床医の判断が診療録を読み返し判断するしかない。そのため、レセプト/DPCなどの診療報酬請求情報を使用した分析、または電子カルテ情報を用いてビッグデータの分析においては、MACEなどのイベントをアウトカムにした研究をすることができないという限界がある。本研究では、電子カルテの記事情報から自然言語処理を活用して自動的にMACEであると判断するためのシステムを開発し、電子カルテ情報を用いたMACEのビッグデータ分析を行うためのシステムを開発する。日本語で記述される電子カルテからの臨床データベースにおいては初めての試み

である。

日本循環器学会の事業で実施している医療コストがかかる疾患・治療（心筋梗塞・狭心症とその病態に対するステント治療、重症心不全とそれに対する再同期療法（CRT））と循環器領域で特にその重要性が指摘されている疾患（急性心不全など）を抽出し、医療の質とその妥当性を検証するため時間軸を念頭においたデータベースである「臨床効果データベース」を用いて自然言語解析による診断判定システムの構築をおこなうことを目的としている。

B. 研究方法

疾患コホート研究であり、虚血性心疾患、急性心不全の患者を対象とする。対象施設は、国立循環器病研究センター、東京大学、自治医科大学、自治医科大学さいたま医療センター、東北大学、九州大学にてデータの収集を行う。臨床効果データベースから、患者基本情報、診断名、入退院情報、経時的な内服薬、経時的な臨床検査情報、経時的な生理検査情報、経時的な心臓カテーテル検査情報を取得する。別途、電子カルテの記事情報を、臨床効果データベースと同じ匿名化番号にて匿名化したIDにて連結可能匿名化して受け取り、臨床効果データベースのデータと連結を行うことにより、電子カルテの記事情報と臨床データの結合を行う。さらに、入退院情報、検査結果等のデータからMACEによる入院かどうかの判断を行う。これらのMACEのそれぞれのイベントに対して、ICD10を用いて病名のコーディングを行う。次に、電子カルテの記事の医療用語を傷病名、愁訴、身体所見、検査、治療のそれぞれに分類を行う。これらの医療用語の出現とMACEの関係を機械学習（サポートベクターマシンやディープラーニング等）の手法を用いて学習を行い、電子カルテ記事からMACEかどうかを判定する予測モデルを作成する。

倫理面への配慮として、協力病院からデータを収集する際に、連結可能匿名化とし、個人識別情報および対応表は施設管理者の保管元、施設外に持ち出さないように厳格に管理する。

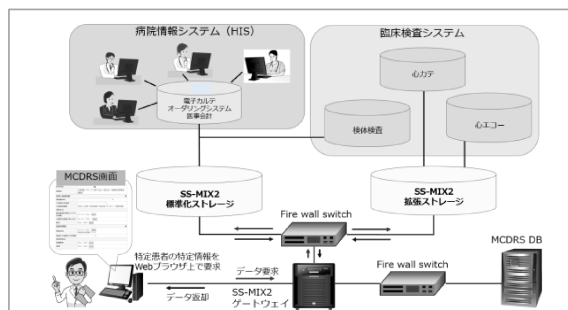
C. 研究結果

本年度は、SS-MIX2 拡張ストレージに心電図、心臓超音波、心臓カテーテル検査結果の標準項目と形式とを定めた日本循環器学会標準出

力フォーマット（Standard Export data format:SEAMAT）を用いて、国立循環器病センター、東北大学、自治医大、自治医大付属さいたま医療センター、東京大学、九州大学のデータを実際に格納する準備を行った。今回過去データも対象とするためCSVファイルからデータ転送するプログラムを開発した。心電図については心電図サーバからSEAMAT形式でSS-MIX2 拡張ストレージに書き出す準備が整った。（図1）

多目的臨床データレポジトリシステム

図1 臨床効果データベース



臨床効果データベースの情報を利用して人工知能に学習をさせる。

（MCDRS）（図2）を用いて症例登録画面の作成を開始した。

図2 多目的臨床データレポジトリシステム
症例登録の画面例（各医療機関の登録者）
電子カルテデータと簡単に連携できる

MCDRS を用いることにより、SS-MIX2 の標準ストレージより患者基本情報、臨床血液検査結果、また、SS-MIX2 拡張ストレージより心電図、心エコー、心臓カテーテル検査結果を抽出してMCDRSにデータを格納することが可能である。

さらに、自然言語処理のためのフレームワークの検討として、形態素解析器の検討および医学用語辞書の精度向上のための検討を行った。申請時には MeCab の利用を想定していたが、京都大学で開発された Juman++ の性能が高いことがわかり、その利用を検討した。実際に用いるためには、形態素解析の精度ばかりでなく、辞書データのメンテナンス性や、解析結果の取得のしやすさ、実装のしやすさ、速度、安定性などが要求される。検討した結果としては今のところ十分に利用可能と考えており、今後、各施設より取得した実カルテ記載データ、特に数値情報等が混在するテキストデータでの検証を行う。また、辞書については急性心筋梗塞、心不全に関する用語のチェックを継続して行っている。病院情報システムからのデータ抽出としては、国立循環器病研究センター病院では N 社の病院情報システムが導入されており、データウェアハウス (DWH) が導入されている。DWH にはいわゆる電子カルテ本体より SOAP 記載が転送されており、これらのデータは抽出プログラムを構築することにより、DWH のデータベースからサンプルデータを抽出して、形態素解析の検証を実施した。

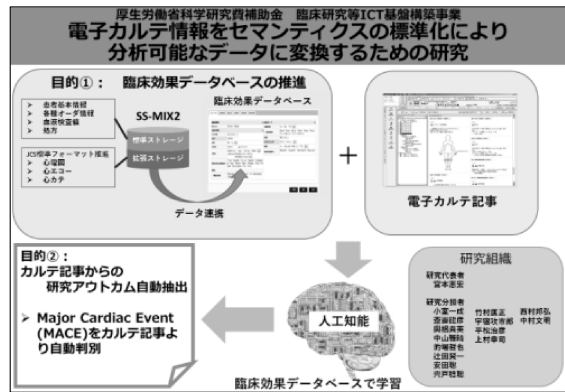
D. 考察

研究により、レセプト/DPC などの診療報酬請求情報を使用した分析、または厚生労働省電子的診療情報交換推進事業による標準化ストレージである SS-MIX2 データなどの電子カルテ情報を用いてビッグデータの分析において、MACE などのイベントをアウトカムにした研究をすることができるようになる。それにより高齢化社会の中で急増する循環器疾患の予後を改善させ医療費を適正化するための医療分析が可能となる。

E. 結論

この研究成果が電子カルテに実装されるようになれば、診療録の記載や各種検査・投薬オーダーを行う際の警告(アラート)の強化や、記載漏れや監査に対応した病名推薦システムを構築することができる。本研究によりレセプト/DPC などの診療報酬請求情報を使用した分析、または電子カルテ情報を用いてビッグデータの分析において、MACE などのイベントをアウトカムにした研究を実施することができるようになる。それにより高齢化社会の中で急増する循環器疾患の予後を改善させ医療費を適正化するための医療分析が可能となる。また、この研究成果が電子カルテに実装されるようになれば、診療録の記載や各種検査・投薬オーダーを行う際の警告(アラート)の強化や、記載漏れや監査に対応した病名推薦システムを構築することができる(図 3)。

図 3



循環器疾患の辞書作りは、電子カルテ記事の自然言語処理を行うのに必要な作業であり、他の研究にも応用活用であり、電子カルテに実装することが可能となる成果物である。心エコー、心臓カテーテル検査レポートを SS-MIX2 から取得できるようになれば、ベンダーの違う電子カルテ同士でもデータの連携が可能となる。

F. 健康危険情報
なし

G. 研究発表

1. 論文発表 なし
2. 学会発表 なし

H. 知的財産権の出願登録状況

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし

平成28年度厚生労働科学研究費補助金
(政策科学総合研究事業(臨床研究等ICT基盤構築研究事業))
分担研究報告書

日本循環器学会データ出力標準フォーマットガイドラインに基づく
心臓超音波レポートおよび心臓カテーテル検査レポートデータベースの作成

研究分担者 中村 文明・西村邦宏・斉藤 能彦・安田 聡・宮本恵宏

研究要旨

英国や米国では電子カルテデータから情報を抽出してデータベース化を行う事業が進んでいる。本邦では電子カルテベンダーごとにデータ保存形式に互換性がなく、施設間を超えて電子カルテデータから同一の形式でデータを抽出することが困難であった。本研究では、日本循環器学会データ出力標準フォーマットガイドライン(Standard Export data format: SEAMAT)に従って、心臓超音波レポートおよび心臓カテーテル検査レポートを作成し、全国でこれらのデータを収集できる仕組みを構築するためのパイロット研究を行った。国立循環器病研究センターのレポートをSEAMAT形式に自動で変換するシステムの構築を試みたが、項目の不足によりSEAMATで規定されたすべての項目を作成することはできず、部位や測定値の自由記載による変換困難例が存在することが判明し、自動での変換から手動での変換に切り替えた。それぞれの施設のレポートをSEAMAT形式に変換することは困難であり、レポートの記載方法も標準化していく必要があると考えられた。

A. 研究目的

英国や米国では電子カルテデータから情報を抽出してデータベース化を行う事業が進んでいる(Denaxas, et al. International Journal of Epidemiology 2012;41:1625-1638)。本邦では電子カルテの普及は早かったものの、ベンダーごとにデータ保存形式に互換性がなく、施設間を超えて電子カルテデータから同一の形式でデータを抽出することが困難であった。日本循環器学会の事業である「臨床効果データベース」は、医療コストがかかる疾患・治療(心筋梗塞・狭心症とその病態に対するステント治療、重症心不全とそれに対する再同期療法(CRT))と循環器領域で特にその重要性が指摘されている疾患(急性心不全など)を抽出し、医療の質とその妥当性を検証するため時間軸を念頭においたデータベースであるが、その特徴としては、厚生労働省の診療情報交換推進事業による標準化ストレージであるSS-MIX2を使用することにより、半自動的にデータ収集を行う点である(http://www.ss-mix.org/cons/ssmix2_about.html)。SS-MIX2形式でデータを保存することにより、異なるベンダーの電子カルテからの情報も同じ形式でデータを抽出することが可能になると考えられる。

しかしながら、循環器疾患の情報として必要不可欠な心電図、心臓超音波レポート、心臓カテーテル検査レポートのデータはSS-MIX2の形式では規定されていない。このような状況を鑑みて、日本循環器学会では日本循環器学会データ出力標準フォーマットガイドライン(Standard Export data format: SEAMAT)を作成し、この形式で保存されたデータをSS-MIX2拡張ストレージに保存することにより、データの収集を行えるようにした(http://www.j-circ.or.jp/itdata/jcs_standard.htm)。

本研究では、SEAMAT形式による心臓超音波レポートおよび心臓カテーテル検査レポートを作成し、全国でこれらのデータを収集できる仕組みを構築するためのパイロット研究を行う。

B. 研究方法

国立循環器病研究センターにおいて、国立循環器病研究センターの心臓超音波レポートおよび心臓カテーテル検査レポートから、SEAMAT形式に自動的に変換するシステムを構築するよう設計した。心臓超音波レポートおよび心臓カテーテル検査レポートのシステムからそれぞれレポートデータをCSV形式で排出し、医療専用で作成されたエンタープライズ・サービス・バスであるHealthShare®を用いてSEAMAT形式に変換し、SS-MIX2拡張ストレージに保存するようにした。

国立循環器病研究センターのレポートのデータ項目を、そのままSEAMAT形式で保存することは不可能であったため、HealthShare®内部で変換することとした。

(倫理面への配慮)

本研究は国立循環器病研究センター内の倫理委員会の承認を得て行った。

C. 研究結果

SEAMATガイドラインには、詳細な設計書が記されておらず、詳細な情報を得るには保健医療情報システム工業界(JAHIS)規約を参照する必要があることが判明した。SEAMAT形式でデータを保存するための詳細な設計書を添付資料に示す。

心臓超音波レポートに関しては、207項目のSEAMAT形式のデータのうち77項目が国立循環器病研究センターのレポートから作成可能であった(添付資料1)。心臓カテーテル検査レポートにおいては、143項目中100項目が作成可能であった(添付資料2)。作成できなかった項目は国立循環器病研究センターのレポートでは記載がされていない項目であった。

また、国立循環器病研究センターのレポートでは、冠血管の部位において自由記載が可能となっているため、様々な表記が存在し、自動変換が困難であることが判明した。心臓カテーテル検査レポートに関しては、SEAMAT形式への自動変換は

困難であると判断し、自由記載部分を人間が判断して変換するようにシステムを構築した。

D. 考察

本研究では、心臓超音波レポートおよび心臓カテーテル検査レポートを自動的にSEAMAT形式に変換して保存することを試みた。SEAMATガイドラインのみではSEAMAT形式のデータを作成することは困難であり、詳細はJAHISの規約を参照しなければならないことが判明した。JAHISの規約のみでは、作成者の判断によって形式が変わりうる可能性がある。本研究では、心臓超音波レポートおよび心臓カテーテル検査レポートの設計書を作成したので、この設計書を基に今後は他の施設がSEAMAT形式への変換を行えば標準化が可能ではないかと考えられる。

一方で、記載項目が足りないため、SEAMATで規定されているすべての項目を満たすことはできなかった。SEAMATで規定されている項目には、日常の臨床で必ずしも記載されている項目以上のものが含まれており、これが規定されている項目をすべて含むことができなかった理由だと考えられる。今後、全国でデータを集めていく場合にはデータの保存形式を標準化するだけでなく、レポートの記載項目で必須とされる最低限のものを循環器学会等で議論し、規定していく必要があるだろうと考えられた。

また、レポートには部位や測定値に幅を持たせたり、参考値として記載したりしている場合があるため、これらを研究用のデータとしてはそのまま収集できないことが判明した。SEAMAT形式でデータを保存する際には、このようなデータの取り扱い方法も標準化する必要があると考えられる。

E. 結論

心臓超音波レポートおよび心臓カテーテル検査レポートをSEAMAT形式に保存するための設計書を作成した。それぞれの施設のレポートをSEAMAT形式に変換することは困難であり、レポートの記載方法も標準化していく必要があると考えられた。

F. 研究発表

1. 論文発表
- 1) Maeda E, Nakamura F, Boivin J, Kobayashi Y, Sugimori H, Saito H. Fertility knowledge and the timing of first childbearing: a cross-sectional study in Japan. *Hum Fertil (Camb)*. 2016 Oct 5:1-7.
- 2) Okamoto M, Nakamura F, Musha T, Kobayashi Y. Association between novel arterial stiffness indices and risk factors of cardiovascular disease. *BMC Cardiovasc Disord*. 2016 Nov 7;16(1):211.
- 3) Okamoto M, Kobayashi Y, Nakamura F, Musha T. Association Between Nonrestorative Sleep and Risk of Diabetes: A Cross-Sectional Study. *Behav Sleep Med*. 2016 May 6:1-8. [Epub ahead of print]
- 4) Maeda E, Nakamura F, Kobayashi Y, Boivin J, Sugimori H, Murata K, Saito H. Effects of fertility education on knowledge, desires

and anxiety among the reproductive-aged population: findings from a randomized controlled trial. *Hum Reprod*. 2016 Sep;31(9):2051-60.

- 5) 坂田 弥生, 森岡 典子, 中村 文明, 豊川 智之, 小林 廉毅. 病院に勤務する看護師の分布とその関連要因の検討. *日本公衆衛生雑誌* (0546-1766)63 巻 7 号 Page367-375(2016.07).

2. 学会発表

- 1) Nakamura F, Nishimura K, Guzman-Castillo M, Sekikawa A, Capewell S, Miyamoto Y, Kuller L, O'Flaherty M. Explaining the fall in coronary mortality in Japan between 1980 and 2012: IMPACT modelling analysis. *Society for Social Medicine 60th Annual Scientific Meeting*, York, UK.

G. 知的財産権の出願・登録状況 (予定を含む。)

1. 特許取得
なし
2. 実用新案登録
なし
3. その他
なし

平成28年度厚生労働科学研究費補助金
(政策科学総合研究事業(臨床研究等ICT基盤構築研究事業))
分担研究報告書

データ転送プログラムによるデータ収集に関する研究

研究分担者 中山 雅晴・興梠 貴英・的場 哲哉

研究要旨

厚労省標準保存形式である SS-MIX2 の拡張ストレージに循環器特有の検査結果を転送し、データを2次活用するための基盤を整備する目的のもと日本循環器学会で標準出力フォーマット(SEAMAT)を定めた。本研究ではその実装を進めるべく、特に csv ファイルから SS-MIX2 ストレージへデータを転送するための仕様を策定したのち、日本循環器学会 IT/DB 委員会からの承認を得て、変換ツールを作成した。そのツールをもとに各施設のサンプルデータを変換し、問題点の洗い出しを行った。また、東北大は各施設に先行して、実際にサーバー上で動かす準備を行った。今後、各施設での問題点の洗い出しと対応が済み次第、ツールの導入とデータ出力を順次展開する。

A. 研究目的

電子カルテ情報を分析可能なデータに変換し、活用するためには必要なデータ群を病院情報システムから該当データベースへ転送する仕組みが必要である。しかしながら病院情報システムはベンダーにより様式が異なり、互換性は担保されていない。そのため、システムベンダーに依存しないデータ保存形式として厚生労働省電子的診療情報交換推進事業による Standardized Structured Medical Information eXchange(SS-MIX)ストレージ形式が定められ、現在 SS-MIX version2 (SS-MIX2) が多くのプロジェクトで用いられている。SS-MIX2 ストレージは標準化ストレージと拡張ストレージとに分けられ、前者には標準化対応の進んだ病名や検査値、処方などのデータが格納される。一方、拡張ストレージにはそれ以外のデータが格納されるため、コード化やデータ形式の標準化が不十分であることからデータの2次利用にはさらなる手順や負荷を要すこととなり、活用は困難な状況となっている。循環器分野においては、心電図、心臓超音波、心臓カテーテル検査などが患者の病態を評価する上で必須となるが、SS-MIX2 上では拡張ストレージに保存されるため、これらのデータを活用するために必要な標準項目名や単位、形式などを決定する必要があった。この状況を鑑み、日本循環器学会では2015年度に SS-MIX2 拡張ストレージへデータを出力するための標準フォーマット Standard Export datA for MAT (SEAMAT) を定めた (http://www.j-circ.or.jp/itdata/jcs_standard.htm)。SEAMAT は保健医療情報システム工業会(JAHIS)標

準規格としても制定され、2015年に先行した安静時心電図検査と心臓用音波検査に追加して、2016年6月には「JAHIS心臓カテーテル検査レポート構造化記述規約Ver.1.0」が追加された。これらの規約では、項目名や単位の他に、国際標準化コードとして LOINC にも対応している。本研究における分担研究では臨床効果データベースの一環として、実際に SEAMAT を用いて患者データを収集することを目的とする。

B. 研究方法

本研究対象施設である国立循環器病センター、東京大学、東北大学、九州大学、自治医大における生理検査システムや心臓カテーテル検査レポートシステムから SEAMAT によるデータ出力を行う。実現するためにはいくつかの方法があり、SEAMAT 対応のシステムを新規購入すること、各システムから SEAMAT 対応によるデータ出力プログラムをカスタマイズすること、csv 形式でデータを出力し、それを SEAMAT 形式に変換して SS-MIX2 に出力することが挙げられる。コストを換算するとが一番安価で汎用性が高いため、分担研究者である、東北大 中山、九州大 的場、自治医大 興梠の担当は のプログラム開発と実装である。

倫理面への配慮

実名情報をストレージで扱う範囲ではそれぞれの施設の院内ルールに従う。ストレージからの情報共有に関しては、本研究に関するそれぞれの施設の倫理委員会の裁定に遵守し、とりわけ個人情報の保護の観点からその取扱いには十分に配慮を行う。

C. 研究結果

日本循環器学会IT/DB委員会でも上述と同様の議論が行われ、CSVファイルからのデータを転送するプログラムを開発することが決定した。複数社の提案のうち、厳選の結果ネクシス社が選定され、平成28年1月に契約が締結された。その技術指導には国際化標準推進組織である Integrating the Healthcare Enterprise (IHE) - 日本の循環器技術委員会が担うことも併せて決定された。

平成28年1月から2月にかけて、各施設から心電図や心臓超音波、心臓カテーテル検査結果より患者名やIDなどを匿名加工化したサンプルデータをネクシス社に送付、課題や疑問点を各施設間で協議し、移行にあたる問題点を順次整理した。予想されたことであるが、各組織間での項目名や定義の違いが大きいため作業量が膨大となっている。さらに、東北大学では作成したツールを用いてxmlファイルに変換、SS-MIX2形式における互換性・正確性・基準への遵守等を確認している。項目の確認を経て、実際にサーバー上で動かす準備も行っている。今後、各施設での問題点の洗い出しと対応が完了し、平成29年3月にツールの導入とデータ出力を行った。

D. 考察

ビッグデータ時代と言われ、かつてないほど膨大なデータを用いて新規の知見を得ることが可能となり、最近では深層学習の進化によりさらなる発展が期待されている。医療分野においても実用が進んでおり、その競争も激化している。一方活用されるべきデータ自体は、テキストや数値、画像を含めた診療情報に限らず、バイタルデータやゲノムデータなど広範囲に渡るリソースが対象となっている。しかしながら、現状では、診療情報の中心はレセプトデータやDPCデータであり、本来主体となるべき診療記録そのものを扱う電子カルテから直結した診療情報の活用は、期待されているほど十分に進んでいない。それは、電子カルテの導入が遅滞し、医療施設数全体としての導入率が30 - 40%程度であるため、多くの情報が依然紙媒体中心であることが理由としてあげられる。さらに、電子カルテであってもベンダー間で互換性がないため、その活用が制限されることも大きな一因として挙げられる。この問題に対しては、厚労省の定める標準保存様式SS-MIX2の浸透により、患者基本情報、病名、処方、注射、検体検査などのデータはベンダーを問わず収集が可能となった。実際、PMDAを中心として薬剤投与による副作用を早期に検出することを目的とした医療情報データベース MID-NET プロジェクト (<http://www.mhlw.go.jp/file/05-Shingikai-11121000-Iyakushokuhinkyoku-Soumuka/0000038207.pdf>) や糖尿病学会が中心となって大規模に糖尿病患者データを収集するJ-DREAMS(<http://jdreams.jp/>) が、

SS-MIX2ストレージからの診療情報を用いた事業として開始されている。

今後、SS-MIX2ストレージ内の情報を用いることが主流となっていくことが予想されるが、現状のSS-MIX2では標準化ストレージ内の一般的なデータが中心であるため、診療科特有のデータが格納される拡張ストレージの活用が大いに望まれるところである。本研究は、循環器領域特有のデータを収集されるために必須の課題であり、進展が早期に望まれる。幸い、日本循環器学会のIT/DB委員会を中心に学会のコンセンサスがとれ、他の関連学会の承認も得ることが出来ている。今後は実装が重要であることが自明であり、本事業を参加施設間で実現させ、さらに早期に国内他施設へ展開させていく。

E. 結論

循環器領域のデータを効率的に収集、2次活用するためにSEAMATに準拠したデータ収集の基盤を整備する。本年度はcsvファイルからのSEAMAT出力をいくつかの施設で実装した。今後他施設へ展開させていく。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表 なし
2. 学会発表
 - 1) 第80回日本循環器学会学術集会 (2016年3月)
口演 Japanese Circulation Society Releases the Standard Export Data Format for Standardized Structured Medical Information eXchange Extended Storage. M.Nakayama et al.
 - 2) 第64回日本心臓病学会学術集会 (2016年9月) シンポジウム・ICTを活用したビッグデータからの知識生成：我国の現状と課題。口演『診療情報ストレージ標準規約SS-MIX2を用いたビッグデータ収集における現状と問題点』中山雅晴
 - 3) 第36回医療情報学連合大会 (2016年11月) IHEチュートリアル セミナー『日本循環器学会標準出力 (SEAMAT) について』中山雅晴
 - 4) 第81回日本循環器学会学術集会 (2017年3月) ポスター Developing a converter program for a standardized data format authorized by the Japanese Circulation Society. M. Nakayama et al.

H. 知的財産権の出願・登録状況

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし

平成28年度厚生労働科学研究費補助金
(政策科学総合研究事業 (臨床研究等 ICT 基盤構築研究事業))
分担研究報告書

臨床効果データベース事業の学会支援に関する研究

分担者研究者 小室一成 東京大学医学部附属病院 循環器内科 教授

研究要旨

電子カルテ記事から必要な情報を自動的に抽出し、臨床効果データベースと統合することにより主要有害心血管イベント (MACE) を判定するシステムを構築する。

A. 研究目的

電子カルテの記事情報から自然言語処理を活用して自動的に Major Adverse Cardiac Event (MACE) とよばれる主要有害心血管イベントを判断するシステムを開発し、電子情報による医療ビッグデータ自動分析を行うためのシステムを構築する。

B. 研究方法

各施設の臨床データフォーマットを標準化して蓄積する SS-MIX2 データベースとデータ抽出用の MCDRS システムを活用して進めていく。

(倫理面への配慮)

東京大学医学部附属病院のカルテ情報の利用にあたって、東京大学医学部倫理委員会の承認を申請中で、今後その指示に従う。データを収集する際に個人情報情報は削除して連結可能匿名化とし、個人識別情報および対応表を施設外に持ち出さないように厳格に管理する。

C. 研究結果

現在、東京大学は、革新的研究開発推進プログラム (ImPACT) の「社会リスクを低減する超ビッグデータプラットフォーム」に参加しており、東京大学循環器内科の心臓カテーテルレポートシステム (CAIRIS-DB および GoodNet) がス

ーズに SS-MIX2 に蓄積され、MCDRS を用いて抽出するシステムが構築されたところである。

循環器領域以外での SS-MIX2、MCDRS の活用状況について、糖尿病領域では「診療録直結型全国糖尿病データベース事業」(J-DREAMS) が、腎臓病領域では「我が国における慢性腎臓病患者に関する臨床効果情報の包括的データベース構築に関する研究」(J-CKD-DB) が、SS-MIX2 および MCDRS を活用している。J-DREAMS は 2016 年 11 月末時点で 3 2 施設が参加している。

D. 考察

今後、SS-MIX2 ストレージ内に蓄積された電子カルテ情報を用いた臨床研究が予想されるが、現状、導入している施設の多くは SS-MIX2 は標準化ストレージ内の一般的なデータが中心であるため、循環器疾患診療に特有のデータが格納される拡張ストレージの活用が必須である。本研究は、日本循環器学会が実施している「臨床効果データベース構築」事業と連携し、参加施設間で実現させ、さらに国内他施設へ展開させていく。

E. 結論

本年度は、心臓カテーテルレポートシステム（CAIRIS-DB および GoodNet）よりレポートデータが SS-MIX2 に蓄積され、虚血性心疾患用 MCDRS を用いて抽出するシステムが構築された。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表 なし
2. 学会発表 なし

H. 知的財産権の出願登録状況

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし

平成28年度厚生労働省科学研究補助金
(政策科学総合研究事業(臨床研究等ICT基盤構築研究事業))
分担研究報告書

電子カルテ情報のSSMIX2連携に関する研究

研究分担者 宇宿 功市郎・辻田 賢一

研究要旨

熊本大学医学部附属病院での電子カルテ情報のSS-MIX2企画での利活用に関する現状について検討を行った。電子カルテ情報はDWHに蓄積され、SSMIX2標準データとして出力可能と確認できた。

A. 研究目的

複数病院が参加する共同研究においては標準化した情報の連携を行い、確実な情報の収集が必要である。今回「臨床研究等ICT基盤構築研究事業」に参加するにあたり、熊本大学医学部附属病院における厚生労働省電子的診療情報交換推進事業による標準化ストレージであるSS-MIX2データの電子カルテ情報の利活用に関する現状について検討を行うこととした。

B. 研究方法

熊本大学医学部附属病院においては、地域医療機関との情報連携において厚生労働省電子的診療情報交換推進事業による標準化ストレージであるSS-MIX2データでの連携を推進してきている。平成29年1月には病院内情報システムの更新を行うこととなり、この際にもSSMIX-2での情報連携を企画している。この状況が今後も行われるかを検討した。

(倫理面への配慮)

情報連携についての院内掲示を行うとともに、連携の際には患者から同意を得ることとしている

C. 研究結果

更新された電子カルテからは、データウェアハウスDWHに情報連携、蓄積されている。このDWHから厚生労働省電子的診療情報交換推進事業による標準化ストレージであるSS-MIX2データの出力が行われていることを確認した。

D. 考察

今回の研究において熊本大学医学部附属病院病院情報システムからの情報は厚生労働省電子的診療情報交換推進事業による標準化ストレージであるSS-MIX2データと連携可能であることが分かった。本研究事業の準備は出来ているものと考えている。

E. 結論

本事業への参加は可能である。

F. 健康危険情報 なし

G. 研究発表

1. 論文発表 なし
2. 学会発表 なし

H. 知的財産権の出願・登録状況

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし

平成28年度厚生労働省科学研究補助金
(政策科学総合研究事業(臨床研究等 ICT 基盤構築研究事業))
分担研究報告書

自然言語解析による診断名判断システムの開発

研究分担者 穴戸 稔聡・平松 治彦・上村 幸司・竹村 匡正

研究要旨：

リアルワールドにおける電子カルテ記事の分析を行うための下準備として、(1) 自然言語処理のためのフレームワーク、および (2) 病院情報システムからのデータ抽出方法について検討を行った。(1)形態素解析器としては、Juman++の有用性が明らかとなった。今後は、収集したカルテ記載データを用いた検証を行い、その結果を医学用語辞書のメンテナンスへフィードバックしていく必要がある。(2) 病院情報システムからのデータ抽出については、電子カルテ記載情報のうち、SOAP 情報は DWH から抽出できることが確認できたが、報告書やサマリは、部門システムごとに抽出する必要があり、抽出方法については引き続き検討を要することがわかった。また、テキスト情報を失った PDF 形式のデータが多く存在していたため、自然言語処理を行う前処理としての OCR などの検証が今後必要であることがわかった。

A. 研究目的

本研究の目的は、電子カルテシステム内に蓄積された所見・報告書・サマリなどのテキスト情報から、自然言語処理および機械学習を用いて、Major Adverse Cardiac Event (MACE) とよばれる主要有害心血管イベントを予測するモデルを構築することである。

テキスト情報を対象に機械学習をするためには、記載内容を自然言語処理し、適切な用語を抽出する必要がある。しかし、精度の良い用語抽出には、最適な自然言語解析器の選択や医学用語辞書のチューニングをしなくてはならない。また、対象とする所見・報告書・サマリなどのテキスト情報は、様々なシステムで作成・保管され、その形式も多様であるため、本研究で利用できる形式で抽出・収集可能か検証が必要である。

そこで、今回は、実際の電子カルテ記事の分析を行うための前段階として、1. 自然言語

処理のためのフレームワークおよび、2. 病院情報システムからのデータ抽出方法について検討を行った。

B. 研究方法

1. 自然言語処理のためのフレームワーク

形態素解析器および医学用語辞書の精度向上のための検討を行った。検討の際には、形態素解析の精度・解析結果の取得のしやすさ・実装のしやすさ・解析速度・解析結果の安定性・医学用語辞書データのメンテナンス性などの観点から評価した。

2. 病院情報システムからのデータ抽出方法

対象となる所見・報告書・サマリなどの情報が、どのシステムにどのような形式で保管されているか確認し、抽出方法を検討した。

(倫理面への配慮)

本研究において診療データを利用する際

には、国立循環器病研究センターなど参加施設の倫理委員会の承認を得てその指示に従う。情報収集協力病院からデータを収集する際には、個人情報には削除して連結可能匿名化とし、個人識別情報および対応表を施設外に持ち出さないように厳格に管理する。

C. 研究結果

1. 自然言語処理のためのフレームワーク

形態素解析器については、申請時には MeCab の利用を想定していたが、本年度に京都大学で開発された Juman++ の性能が高いことがわかり、本システムの利用を検討した。その結果、解析器の精度・利便性などの面から十分に利用可能と考えられた。

また、医学用語辞書については、継続的に手作業でメンテナンスを行い、医学分野における形態素、構文情報、シソーラスなどの整備を行った。

2. 病院情報システムからのデータ抽出方法

電子カルテに記載されている所見(いわゆる SOAP 情報)は、データウェアハウス(DWH)のデータベースから抽出可能なことが確認できた。

一方、放射線・心エコー・心電図などの各種検査報告書は、各部門のレポートシステムで記載・保管された後、PDF 化されて電子カルテシステムから参照できる形で保管されていた。

同様に、診療情報提供書や退院サマリも専用の文書作成システムにおいて作成され、印刷・押印・スキャンを経て、PDF 形式で保管されていた。

また、手術記録は、作成方法が統一されておらず、Word や Excel など記載・印刷されたものをスキャンされて PDF 形式で保管されていることがわかった。

このように、記載情報の多くが、テキスト情報を失った PDF 形式で保管されていること

が確認できた。

D. 考察

1. 自然言語処理のためのフレームワーク

JUMAN++ は、言語モデルとして Recurrent Neural Network Language Model (RNNLM) を用いることにより、単語の並びの意味的な自然さを考慮することが可能な解析器である。そのため、JUMAN、MeCab に比べ大きく性能が向上している。今後、取得された実際の電子カルテ記載データ、特に検査値などの数値情報が混在するテキストデータでの検証を行い、解析精度などの検証を行う予定である。また、その結果を踏まえて、医学用語辞書のメンテナンスを手作業で継続して行う。

2. 病院情報システムからのデータ抽出方法

電子カルテ記事のうち、SOAP 情報は DWH のデータベースから抽出可能なことが確認できたが、その他の情報は、多くがテキスト情報を失った PDF として保管されており、そのままでは自然言語解析を行うことができないことがわかった。

報告書やサマリなどは、各々を作成した部門システムのデータベースから直接抽出する必要があるが、その際には、日常業務に差し障りのないように、システム負荷や時間帯などを考慮した安定的な取得方法を検討しなければならない。

また、手術記録はスキャンされた PDF ファイルしか存在しなかった。したがって、自然言語処理を行うためには、前処理として OCR 等でテキスト情報への変換が必要である。その場合、OCR の精度が問題となるため、今後十分な検証を要する。

E. 結論

本年度は、実際の電子カルテ記事の分析を行うための下準備として、(1) 自然言語処理のためのフレームワークおよび (2) 病院情

報システムからのデータ抽出方法について検討を行った。形態素解析器としては、Juman++の有用性を評価することができた。今後は、収集したカルテ記載データを用いた検証を行い、その結果を医学用語辞書のメンテナンスへフィードバックしていく必要がある。電子カルテ記載情報のうち、SOAP 情報は DWH から抽出できることが確認できたが、報告書やサマリは、部門システムごとに抽出する必要があり、抽出方法については引き続き検討を要する。また、テキスト情報を失った PDF 形式のデータが多く存在していたため、自然言語処理を行う前処理としての OCR などの検証も必要であることがわかった。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

なし

2. 学会発表

なし

H. 知的財産権の出願・登録状況(予定を含む)

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

研究成果の刊行に関する一覧表

書籍

著者氏名	論文タイトル名	書籍全体の 編集者名	書 籍 名	出版社名	出版地	出版年	ページ
なし							

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
なし					