

平成 30 年度厚生労働科学研究補助金

(臨床研究等 ICT 基盤構築・人工知能研究事業) 分担研究報告書

SS-MIX2 分析用データセットおよびシステムモジュールの 作成・開発について

堀口 裕正 国立病院機構本部総合研究センター 診療情報分析部 副部長
岡田 千春 国立病院機構本部総合研究センター 企画役

研究要旨

本分担研究において、国立病院機構本部との調整を中心とした基盤構築を行った。まず、NDCA データの研究利用に向け、倫理審査申請に加えて、内部規定にて定められている内部委員会の調整を図った。また、閲覧・解析に特化した自然言語処理用の研究基盤の構築を行った。

また、NCDA データセットから、そのデータ仕様に基づいた匿名化モジュールの開発を行った。本年度、基本 4 情報を含む単独で個人情報とみなされる情報を削除するモジュールを開発し、そのモジュールを通過させた後に研究者に提供をおこなった。

また、実運用を想定し、SS-MIX の拡張ストレージから医師記録データを読み込み、それをサマリフォーマットに変換して次の処理に渡すことのできるモジュールの開発を行った。

A. 目的

本研究は、電子カルテに実装可能な医療用語の標準化を行うシステムの開発に向け、そのステップとして、臨床現場からのニーズが極めて大きい退院サマリの自動生成技術の実現を目標と定める。電子カルテの自動解析は技術的な難易度が高く、実用的な精度を実現するためには多額の研究開発投資が求められる。そこで、本研究提案では、医療現場に直接的なメリットが生じる研究課題に取り組むことによって、現場の協力と今後の追加的な研究開発投資を呼び込み、その過程を通じて実用性の高い電子カルテの自動解析技術を実現する戦略を採る。初年度、我々が今まで模擬カルテを用いて研究開発を進めてきた標準化技術を、国立病院機構の有する広域電子カルテ網(NCDA)上の実カルテへと適用し、技術的な課題を抽出する。2年目には、NCDAを用いて集積した電子カルテに加えて、退院サマリ情報を用いることで、電子カルテの自動要約技術の検討を行う。3年目においては、両技術の統合により、継続的な精度向上の体制を実現するとともに、研究成果を既存の各社電子カルテへと組み込む枠組みを構築する。本研究により、退院サマリの自動要約技術や紹介状の作成支援技術等、医療用の自然言語処理に関連する多彩な応用技術が実現する。これは、医療現場における負担軽減策として極めて効果が期待される。また、こうした応用の発展により、要素技術である電子カルテ上の記載からの自動情報抽出において、継続的な精度向上が実現する。この手法は、電子カルテにおける用語の標準化技

術単独に研究開発投資を行うことと比して、投資効率が極めて高いと考えられる。さらに、こうして医療用自然言語処理技術が発展することにより、大量の電子カルテからの効率的な情報抽出が実現する。これは健康医療政策に資する統計データの収集コストを劇的に低廉化し、今後、政策に求められる様々なエビデンスを継続的に生み出していく基盤となることが期待される。

なお本分担研究では、本研究における「大規模に電子カルテデータを入手できる体制」として、全国の国立病院 63 施設より年間 190 万患者の電子カルテ情報を自動収集する診療情報集積基盤(NCDA)を構築し、運用している基盤を用い、本研究目的のためのデータの収集・分析活動を行うためのシステム構築及び運用を行うことを1つ目の目的とする。また、各社電子カルテへと組み込む枠組みを構築するためのモジュール開発を行うことを2つ目の目的とする。

B. 方法

国立病院機構本部との調整を中心とした基盤構築を行った。まず、NCDAデータの研究利用に向け、倫理審査申請に加えて、内部規定にて定められている内部委員会での調整結果を踏まえてデータ抽出機能の調整を行い、閲覧・解析に特化した自然言語処理用の研究基盤の構築を行うこととした。その上で、本研究を実施するのに不可欠なNCDAデータセットから、そのデータ仕様に基づいた匿名化モジュールの開発を行い、運用を行うこととする。

さらに、各社電子カルテへと組み込む枠組みを構築するためのモジュール開発として

NCDA が各電子カルテベンダーに要求している SS-MIX 拡張ストレージへの医師記録の書き出しを起点として情報を取得し、処理後同じく NCDA が各電子カルテベンダーに要求している SS-MIX 拡張ストレージへの退院サマリの格納場所に要求するフォーマットで書き戻すシステムを想定し、それらに必要な病院内に必要なモジュールの開発を行った。

C.結果

I 研究用データセットの作成・運用について

国立病院機構が平成 27 年度に構築した NCDA データベースは、平成 30 年度末現在 63 病院が参加、約 27000 床、年間実患者数約 190 万人のデータベースであり、診療日翌日には本部のデータベースに検査値や投薬の情報を含む診療データが届くことになっている。

また、本年度運用 3 年目を迎え、今後、MIA のデータベースで今まで実践してきた分析調査を代替できるポテンシャルを持っている。(参考資料に詳細を添付している)

データベースについて

【国立病院機構 診療情報集積基盤】

(コクリツビョウインキョウ シンリョウジョウホウシュウセキキバン)

英文表記 NHO Clinical Data Archives

省略形の記載法 「NCDA」

省略形の呼称 「クリニカルアーカイブス」

41病院で来院患者ベース 94万人/年 17,800床のデータベース

MIX2 の標準仕様に含まれていないが、JAHIS の提供している仕様を参考に、資料 1 及び 2 で示した仕様で NCDA 内に実装することとした。

本研究はカルテの非定型の記載欄に記入されたデータを使うという研究であり、患者の不利益等を防止するために倫理的な配慮をした上で、倫理審査を受けなければならない。平成 29 年 1 月に国立病院機構中央倫理委員会に侵襲・介入なしの観察研究として倫理審査の申請を行い、3 月に承認された。倫理審査の承認後、データ利用に際して必要な国立病院機構内のデータベース利活用審査委員会への利活用申請を行い、3 月にその承認も受けた。倫理審査申請書については資料 3 に示す。

なお、NCDA データは国立病院機構が契約するデータセンター内で厳重に管理されている。研究に際しては、このデータベースから研究テーマごとに匿名化したサブセットを切り出し、国立病院機構本部内のオンサイト利用に限っている。以上により、データセットの利用対象と利用目的を厳しく制限することにより、患者個人情報の保護を行っている。

また、NCDA データセットから、そのデータ仕様に基づいた匿名化モジュールの開発を行った。本年度、基本 4 情報を含む単独で個人情報とみなされる情報を削除するモジュールを開発し、そのモジュールを通過させた後に他分担研究者に提供出来るようになった。

II 病院での運用をイメージしたシステムモジュールの開発

また、本研究で中心的に使われる医師記録等(経過記録・退院サマリ)については、SS-

本研究課題は研究の公募段階から全国の病院であまねく運用可能なシステムの構築を求められていた。この分担研究においては上記の条件を以下の方法で実現することとし、その実現に必要なモジュールを開発した。

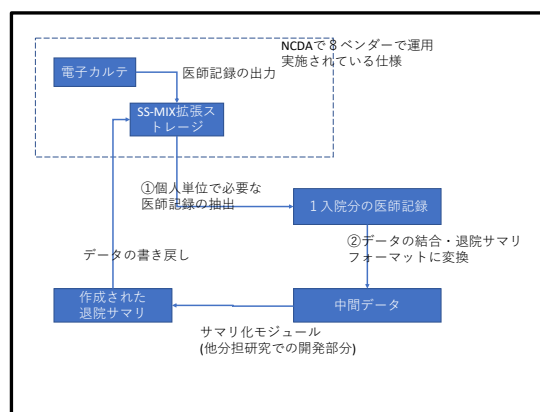
(1) 全国の病院のシステムの接続方式について

本研究では全病院で運用可能な接続として、SS-MIX の拡張ストレージをキーとした運用を採用することとした。

前節で説明をした NCDA における医師記録等の拡張ストレージに対する出力は JAHIS の規約通りである・すでに 8 ベンダー63 病院で実装済みで、現時点でもその病院数を NCDA 内で増加させることに大きな問題を起こしていない。よってこの SS-MIX を利用する方式は研究の公募段階で要求されていた仕様を満たすものと考えられる。

(2) 本研究で開発を行ったモジュールについて

本研究にて想定したモジュール群の概念図を図に示す。この図の形でシステムを実現するものとし、その中の①SS-MIX 拡張ストレージから必要な医師記録データを抽出するモジュール、②抽出したデータを退院サマリフォーマットに変換・結合するモジュールについて開発を行った。



実際に開発するシステム内で組み込みが容易なように実装は Windows 環境下においてコマンドで動作するものとし、go にて開発を行っている。

それぞれのコマンドの Usage は以下の通りである

① のモジュール

ssmix2extf pcopy

Usage:

ssmix2extf pcopy [flags] nhoid

Flags:

-u, --duration string
duration for search condition

-e, --encoding string
encoding of logging format (json | console)
(default "json")

-h, --help help
for pcopy

-d, --logfiledirectory string the
logfile directory

-l, --logfilefilename string the

logfile name (default "pcopy.log")
-c, --lsoutput string csv file
which contain copied files (default
"lsoutput.csv")
-o, --outputdir string output
directory (default "pcopyout")
-p, --patientidlistfile string the file
which contain a list of patient id
-v, --verbose
console verbose mode

② のモジュール

ssmix2extf summary

Usage:

ssmix2extf summary [flags] input.csv

Flags:

-h, --help help for
summary
-o, --outputdir string output
directory (default "summaryout")

モジュールのバイナリおよびソースコード
については <https://github.com/nhoHQ> 内
で公表予定である。

E. 結論

本年度、本研究の他の分担研究を支える分析基盤をきちんと整備・安定した運用を実施することができた。また、本研究課題は研究の公募段階から全国の病院であまねく運用可能なシステムの構築を求められてい他部分について、その実現に必要なモジュールを開発した。

資料1 NCDAにおける医師記録等の仕様書

趣旨

本事業では、各社の **SS-MIX2** モジュールの拡張ストレージへの出力機能を利用し、以下の情報を出力することを求めている。その際、**SS-MIX2** 拡張ストレージ構成の説明と構築ガイドライン Ver.1.2d (以下、ガイドライン) に記載している仕様に対応していること。また、トランザクションストレージ、インデックスデータベースも同時に生成すること。

- 経過記録
- 退院時サマリー
- 診療情報提供書

以下に仕様を示す。

ドキュメントデータ 物理構造

```
|-- 拡張ストレージ ルートフォルダ
  |-- 患者 ID 先頭 3 文字
    |-- 患者 ID 4~6 文字
      |-- 患者 ID
        |-- 診療日
          |-- データ種別
            |-- コンテンツフォルダ
              |-- 主文書ファイル
```

診療日

特に指定しない。

データ種別

ガイドライン P4 (4) 「データ種別フォルダ」について に則ること。

```
[ローカル文書コード]^ローカル文書名称^[ローカルコード体系コード]^標準文書コード^
標準文書名称^標準コード体系コード
```

以下のように標準コードに対しローカルコードが複数あることは許容される。

L12345^入院診療録^99ZZZ^11506-3^経過記録^LN

L12346^外来診療録^99ZZZ^11506-3^経過記録^LN

コンテンツフォルダ

ガイドライン Ver.1.2d P5 (5)「コンテンツフォルダ」について に則ること。

患者 ID_診療日_データ種別コード_特定キー_発生日時_診療科コード_コンディションフ
ラグ

いずれの文書も削除は想定していないが、電子カルテシステムによっては修正はあり得ると考える。その場合、ガイドライン P6 ④修正が発生する場合 に則り改版すること。

主文書ファイル

XML CDA R2 で出力すること。XML ファイル以外に画像ファイルや CSS ファイル等を出
力してもかまわない。

HEADER 部

いずれの文書も JAHIS 診療文書構造化記述規約 共通編 Ver.1.0 に則ること。

P27 6.3.11.検査・診療等行為 "documentationOf/ServiceEvent" によると、
documentationOf の制約・多重度は 0..1 となっているが、経過記録、退院時サマリにつ
いてはこれを 1..1 と読み替えること。

経過記録は serviceEvent classCode(サービスイベントクラスコード)を ENC(診察)とし、
effectiveTime(実施日)は low value、high value とともに記録タイミングを出力すること。

退院時サマリは serviceEvent classCode(サービスイベントクラスコード)を ACCM(入院、
滞在)とし、effectiveTime(実施日)は low value に入院タイミング、high value に退院タイミ
ングを出力すること。

タイミングの粒度は日以上であれば良い。

BODY 部

診療情報提供書は、日本 HL7 協会 患者診療情報提供書 規格 Ver.1.00 に則ること。

診療情報提供書以外は、XML の文法に則ること

2. NCDA システム仕様書

SS-MIX2 を用いた診療情報データベース構築の為の SS-MIX2 モジュール技術仕様書

1. システム要件

国立病院機構の各病院にて「国立病院機構診療情報分析基盤(NCDA)」に参加する為に調達する SS-MIX2 モジュールの機能は以下の通りである。但し、本体の電子カルテシステム等の仕様上、作成が不可能であるものについては作成を要しない。その場合、何が不可能かを導入標準作業手順書に記載すること。

1.1 SS-MIX2 Ver.1.2d 機能

SS-MIX2 Ver.1.2d に準拠することとして、以下の機能を有すること。

- 日本医療情報学会発行の「SS-MIX2 標準化ストレージ構成の説明と構築ガイドライン Ver.1.2d」、「SS-MIX2 拡張ストレージ構成の説明と構築ガイドライン Ver.1.2d」、「SS-MIX2 標準化ストレージ仕様書 Ver.1.2d」、「標準化ストレージ仕様書別紙：コード表 Ver.1.2d」、「SS-MIX2 拡張ストレージ構成の説明と構築ガイドライン Ver.1.2d 別紙：標準文書コード表」に記載している仕様に対応していること。(尚、当初 Ver.1.2c 準拠としていたが、標準ストレージ部分では Ver.1.2c からの変更点について影響がないため Ver.1.2d 準拠ということとした。)
- 標準化ストレージ、拡張ストレージ、トランザクションストレージ、インデックスデータベースの4つのファイルを生成すること。
- 標準化ストレージにはデータ種別として 36 種のデータを出力すること。

(表 1-1 標準化ストレージ格納データ)

No	データ種別	種別名称	HL7 メッセージ型
1	ADT-00	患者基本情報の更新	ADT^A08

No	データ種別	種別名称	HL7 メッセージ型
2	ADT-00	患者基本情報の削除	ADT^A23
3	ADT-01	担当医の変更	ADT^A54
4	ADT-01	担当医の取消	ADT^A55
5	ADT-12	外来診察の受付	ADT^A04
6	ADT-21	入院予定	ADT^A14
7	ADT-21	入院予定の取消	ADT^A27
8	ADT-22	入院実施	ADT^A01
9	ADT-22	入院実施の取消	ADT^A11
10	ADT-31	外出泊実施	ADT^A21
11	ADT-31	外出泊実施の取消	ADT^A52
12	ADT-32	外出泊帰院実施	ADT^A22
13	ADT-32	外出泊帰院実施の取消	ADT^A53
14	ADT-41	転科・転棟(転室・転床)予定	ADT^A15
15	ADT-41	転科・転棟(転室・転床)予定の取消	ADT^A26

No	データ種別	種別名称	HL7 メッセージ型
16	ADT-42	転科・転棟(転室・転床)実施	ADT^A02
17	ADT-42	転科・転棟(転室・転床)実施の取消	ADT^A12
18	ADT-51	退院予定	ADT^A16
19	ADT-51	退院予定の取消	ADT^A25
20	ADT-52	退院実施	ADT^A03
21	ADT-52	退院実施の取消	ADT^A13
22	ADT-61	アレルギー情報の登録／更新	ADT^A60
23	PPR-01	病名（歴）情報の登録／更新	PPR^ZD1
24	OMD	食事オーダー	OMD^O03
25	OMP-01	処方オーダー	RDE^O11
26	OMP-11	処方実施通知	RAS^O17
27	OMP-02	注射オーダー	RDE^O11
28	OMP-12	注射実施通知	RAS^O17
29	OML-01	検体検査オーダー	OML^O33

No	データ種別	種別名称	HL7 メッセージ型
30	OML-11	検体検査結果通知	OUL^R22
31	OMG-01	放射線検査オーダー	OMG^O19
32	OMG-11	放射線検査の実施通知	OMI^Z23
33	OMG-02	内視鏡検査オーダー	OMG^O19
34	OMG-12	内視鏡検査の実施通知	OMI^Z23
35	OMG-03	生理検査オーダー	OMG^O19
36	OMG-13	生理検査結果通知	ORU^R01

「SS-MIX2 標準化ストレージ構成の説明と構築ガイドライン Ver.1.2d p11」

1.2 拡張ストレージへの出力機能

現在の SS-MIX2 モジュールでオプションとして既に導入している拡張ストレージへの出力機能は、そのまま提供すること。また、1.3.0 で規定する出力を行うこと。

1.3 NHO 対応としての設定

1.3.0 拡張ストレージへの出力機能

各社の SS-MIX2 モジュールの拡張ストレージへの出力機能を利用し、以下の情報を出力すること。その際、日本医療情報学会発行の「SS-MIX2 拡張ストレージ構成の説明と構築ガイドライン Ver.1.2d」に記載している仕様に対応していること。また、トランザクションストレージ、インデックスデータベースも同時に生成すること。

No	データ種別	種別名称	HL7 メッセージ型
1	L-OBSERVATIONS^OBSERVATIONS^99ZL01	バイタル検査結果	HL7 V2.5 ORU^R30
2	^(ローカル名称) ^^11506-3^経過記録^LN	診療録 (外来/入院含む)	HL7 CDA R2
2.1	^(ローカル名称) ^^34108-1^外来診療録^LN	診療録 (外来) (入院・外来が別の場合)	HL7 CDA R2
2.2	^(ローカル名称) ^^34112-3^入院診療録^LN	診療録 (入院) (入院・外来が別の場合)	HL7 CDA R2
3	^(ローカル名称) ^^18842-5^退院時サマリー^LN	退院時サマリー	HL7 CDA R2
4	^(ローカル名称) ^^57133-1^紹介状^LN	診療情報提供書	HL7 CDA R2

1.3.1 バイタル検査結果通知の出力

(1) バイタル検査結果通知のデータを、別紙の形式で拡張ストレージに出力する。尚、「診療日」に出力する日付は **OBX-14** トランザクション日時 (測定した日) とする。

(2) ファイル作成の単位は、データの格納構造として日付の下にあるため、最大でも一日分が1ファイルにまとまっている形とする。一日の中で測定のたびに作成するのでも良い。一日1ファイルなら、特定キーは測定日を出力する。一日に複数回のデータを出力する場合は、特定キーに測定日の時間まで (YYYYMMDDHH) 出力すること。

1.3.2 バイタルデータの項目及び形式等

(1) バイタルデータとして取得する項目は、「拡張期血圧、収縮期血圧、脈拍数、呼吸数、体温」の5項目とする。

(2) **OBX-3** 検査項目に出力するコードは **JLAC10** コードとする。バイタルデータを参考に適切な **JLAC10** を選択すること。

(3) 上記以外の項目を **SS-MIX2** に出力することは問題ないが、今回の対応では扱わない。但し、今後の検討で仕様として扱うことになる場合は、**JLAC10** コードを基準とした標準コードを必須とすることを想定している。この今後想定される検査項目は別表として提供する。

1.3.3 標準コード変換機能

SS-MIX2 データの出力に際しては、コードのマッピング表などに従って、院内のローカルコードを厚労省が定める標準コードに変換する機能を有すること。またマッピング表については、容易にその内容を変更できるマスターメンテナンスプログラム等の機能を有すること。

JLAC10 コード、**JANIS** コード、**HOT** コードについては、機構病院が **NCDA** 事業に参加する場合においては機構から提供する。

1.3.4 標準化ストレージにおける文字コードについて

メッセージの文字コードについては、「標準化ストレージガイドライン」で示されているとおり、1バイト系文字は **ISO IR-6 (ASCII)**、2バイト系文字は **ISO IR87 (JIS X 0208 第一水準、第二水準)** とする。ただし現実には上記以外の文字コードが電子カルテシステムに登録されている可能性があるため、以下のように対応することとする。

- 1 半角カナ文字 → 全角カナ文字に置き換えて **SS-MIX2** に出力する。
- 2 外字 → ■で置き換えて **SS-MIX2** に出力する。
- 3 環境依存文字については変換表を機構より提供するのでそれにより変換して **SS-MIX2** に出力する。

1.3.5 単位の文字表記の統一

SS-MIX2 データの出力に際して、臨床検査データの OBX セグメントの 6 フィールド目の単位の文字表記を統一すること。

【単位の文字表記の統一ルール例】 ASCII コードで表記すること

- ・ かける → . (ドット)
- ・ 乗 → * (アスタリスク)
- ・ μ → u (小文字ユー)
- ・ 語尾に名称 → () で
- ・ $^{\circ}\text{C}$ → cel
- ・ $\%$ → permil
- ・ 個 → pcs

【上記ルールの適用例】

- ・ mL → mL (ASCII コード)
- ・ $X10^2/\mu\text{l}$ → .10*2/uL (かける、乗、 μ)
- ・ /HPF → /(hpf) (語尾に名称)

1.3.6 単位変換機能

SS-MIX2 データの出力に際して臨床検査データの単位に関しては、JLAC10 コードごとに、機構が定める単位に変換を行った上で SS-MIX2 データを生成すること。尚、JLAC10 コード別の単位表は別途機構から提供する。単位表は「SS-MIX2 標準化ストレージ仕様書 Ver.1.2」にも別表として添付する。

【単位変換例】

JLAC10 コード	数値	単位	→	JLAC10 コード	数値	単位
1A025000000127201	10.5	mg/l	→	1A025000000127201	1.05	mg/dL

1.3.7 計測値等の表記方法について

(1) 定性値・検出限界以下・検出限界以上の表記

- OBX（検体検査結果）セグメントの5フィールド目（検査値）に検査結果を記述する場合、現在そのデータ形式はOBX-2フィールドの説明にあるようにNM型、ST型、CWE型のうちいずれかの形式で記述することとなっている。
- 今回の仕様では、定性値・検出限界以下・検出限界以上のデータについては、SN型の表現方法を用いてSN型の“A”を” “（スペース）に置き換える。
- この件の説明は、「SS-MIX2 標準化ストレージ仕様書 Ver.1.2」 P104 表 3-77 検査結果セグメント（OBX）定義 のOBX-2の項目説明にも記述する。

(2) 複数の要素が一つの値で表現されている場合の表記

複数の要素が組み合わせられ一つの結果値として表記されている場合は、それぞれの要素に分離して表記すること。例えば定量値とクラス値が組み合わせられた結果値については、定量値とクラス値に分離する。

【定量値とクラス値の分離の例】

定量値とクラス値が組み合わせられた例

検査名称	院内コード	結果値
ムンプス Virus IgG	001591	2.3(±)
↓		
定量値とクラス値を分離した例		

SS-MIX2 標準コード	院内コード	結果値	備考
5F432143102302304	001591	2.3	
5F432143102302311	001591	+-	(半角スペース2つプラスマイナス)

1.3.8 トランザクションストレージのデータ保持期間

トランザクションストレージのデータ保持期間は、現在の標準化ストレージ及び拡張ストレージを作っているデータの再現に必要な分だけ保持しておくこと。

1.3.9 ST 型の長さ

- RXE-23(与薬速度)は ST 型で長さが 6 であるが、正負の記号と小数点を考慮し(例: +266.865)、本事業では 8 桁まで許容するものとする。
- CX 型は先頭成分が ST 型で長さが 15 であるが、IN1-10(被保険者グループ雇用者 ID)に長い名称の保険者が出力される場合などを考慮し、本事業では CX 型の先頭成分は 30 桁まで許容するものとする。
- XAD 型は第 8 成分(その他地理表示)が ST 型で長さが 50 であるが、全角 50 文字(100 バイト)と解釈しているシステムがあり半角文字で 100 文字登録出来るため、本事業では XAD 型の第 8 成分は 100 桁まで許容するものとする。

1.3.10 トランザクションストレージのファイル切り替え機能

SS-MIX2 の仕様上、トランザクションストレージはカレントの日付が変わった時点、もしくは記録中のトランザクションデータファイルのファイルサイズが一定量を超えた時点で、新たなファイルを作成して記録先を切り替えるものとなっているが、同一日付内において一定時刻(例えば 17:00)を経過した時点で記録先を切り替える機能を追加する。

3. 倫理審査における計画書

**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**

研究責任者：堀口 裕正

独立行政法人国立病院機構本部 総合研究センター
診療情報分析部 副部長

事務局/研究主催

独立行政法人国立病院機構本部 総合研究センター
診療情報分析部

堀口 水本

〒152-8621 目黒区東が丘 2 - 5 - 21

TEL: 03-5712-5133

FAX: 03-5712-5134

E-Mail : horiguchi-hiromasa@hosp.go.jp

第 1.0 版：2017 年 1 月 18 日

1. 背景

本研究では、電子カルテに実装可能な医療用語の標準化を行うシステムの開発に向け、そのステップとして、臨床現場からのニーズが極めて大きい**退院サマリの自動生成技術の実現**を目的とする。これは用語の標準化を目的とする研究として遠回りな課題設定である。しかし、電子カルテの自動解析は技術的な難易度が高く、実用的な精度を実現するためには多額の研究開発投資が求められる。そこで、本研究提案では、医療現場に直接的なメリットが生じる研究課題に取り組むことによって、現場の協力と今後の追加的な研究開発投資を呼び込み、その過程を通じて実用性の高い電子カルテの自動解析技術を実現する。

2. 目的

本研究は、電子カルテに実装可能な医療用語の標準化を行うシステムの開発に向け、そのステップとして、臨床現場からのニーズが極めて大きい**退院サマリの自動生成技術の実現**を目的とする

3. 研究方法

3-1. 研究実施場所

研究実施場所は、国立病院機構本部総合研究センター診療情報分析部（以下、診療情報分析部）研究室及び本部内分析室並びに静岡大学情報学部行動情報学科狩野研究室、岡山大学大学院医歯薬学総合研究科クリニカルバイオバンクネットワークワーキング事業化研究講座研究室、国立保健医療科学院研究情報支援研究センター研究室とする。

3-2. 研究実施期間

研究実施期間は、倫理審査委員会承認後より 2020 年 3 月 31 日までとする。

3-3. 研究対象医療機関と対象患者

研究対象医療機関は、国立病院機構病院に所属する DPC 病院のうち、診療情報集積基盤（以下、NCDA）を運用しデータ提供を行う医療機関とする。

対象患者は2016年1月1日から2019年12月31日までに入院し、退院時サマリを作成した全患者とする。

3-4. 対象データ

研究に用いるデータは、研究対象医療機関より診療情報分析部に提供されたDPCデータおよびレセプトデータ、ならびにSS-MIX2ストレージに格納された情報から抽出した医師記録、退院サマリおよび入院中の検査結果、食事内容および処方内容である。

3-5. 分析方法

(1) 対象

退院サマリを作成した全患者

(2) アウトカム

入院中に記載/記録された情報から退院サマリを自動生成する技術を開発すること

(3) 抽出する項目

入院中の医師記録・退院サマリ・入院中の検査結果、食事内容および処方内容

(4) 解析方法

入院中に記載/記録された情報を元データに、機械学習により自動的に情報収集を行い、退院サマリを自動で作成する。その作成結果と、実際の医師の書いた退院時サマリを比較/検討を行い、自動作成技術の能力評価を行い、またその能力の改善を行っていく。

4. 倫理的配慮

本研究は、ヘルシンキ宣言、人を対象とする医学系研究に関する倫理指針（以下、倫理指針）に基づいて実施する。

4-1. インフォームド・コンセント

本研究は既存試料・情報を用いて実施し、人体から取得された試料は用いな

い。研究対象者等からインフォームド・コンセントは受けないが、倫理指針「第12の1(2)イ」に則り、本計画書の4-3に記す通り、利用目的を含む本研究についての情報を研究対象者等に公開し、研究が実施されることについて研究対象者が拒否できる機会を保障する。なお、NCDA運用による診療情報の蓄積・利活用についての説明及び同意は、各施設での掲示で既に行われている。

4-2. データ管理、個人情報等の取り扱いに関する配慮

研究の実施並びに種々のデータの収集及び取り扱いにおいては、国立病院機構診療情報データベース利活用規程に従うとともに、患者情報の機密保持に充分留意する。

本研究で用いるデータは、研究対象医療機関に2016年1月1日から2019年12月31日までに退院サマリを作成した全患者のデータであり、個人情報等を取り扱う。倫理指針「第15の2(1)」及び国立病院機構診療情報データベース利活用規程に則り、保有する個人情報等について、漏えい、滅失又はき損の防止その他の安全管理のため、下記の措置を講じる。

データは研究対象医療機関で収集され、本部IT推進部に提出される。データが保管されるサーバーを国立病院機構本部2階のセキュリティルームに設置し、セキュリティルーム内でIT推進部システム開発専門職が匿名化処理を行う。研究者は匿名化後のデータを用いて本部内分析室において分析を実施する。

保有する個人情報に関する事項の公表等については、倫理指針「第12の1(2)イ」、「第16の1(1)」及び国立病院機構診療情報データベース利活用規程第6条第3項に則り、個人情報の取扱いを含む研究の実施についての情報を研究対象者等に公開する。

4-3. 本研究における情報公開

本研究では、倫理審査委員会承認後、倫理指針「第12の1(2)イ」、「第16の1(1)」及び国立病院機構診療情報データベース利活用規程第6条第3項に則り、本部ホームページにおいて、本研究の意義、目的及び方法、研究機関、保有する個人情報に関して利用目的の通知、開示、訂正等又は利用停止の求めに応じる手続き並びに保有する個人情報に関する問い合わせや苦情等の窓口の連

絡先に関する情報を公開する（公表する情報については別添資料を参照）。

4-4. 研究成果の公表

本研究の成果は、報告書で公表するとともに、学会・論文で発表する。また、本研究結果を内包したソフトウェアの公表を実施する。データの集計・分析結果については、集団を記述した数値データもしくは機械学習の学習結果データとし、個人が同定されるデータの公表は行わない。

5. 研究経費

本研究は、厚生労働科学研究費補助金（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」（代表 堀口裕正）を用いて研究を実施する

6. 研究組織

総合研究センター診療情報分析部が主体となり、本部医療部、保険医療科学院、静岡大学、岡山大学等から協力を得て、研究を行う。

【研究代表者】

国立病院機構本部総合研究センター診療情報分析部 副部長 堀口 裕正

【共同研究者】

国立病院機構本部 企画役 岡田 千春

静岡大学情報学部行動情報学科 准教授 狩野 芳伸

岡山大学大学院医歯薬学総合研究科

クリニカルバイオバンクネットワーク

事業化研究講座研究室 准教授 森田 瑞樹

国立保健医療科学院研究情報支援研究センター

特命上席主任研究官 奥村 貴史

別添

「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」研究実施に関するお知らせ