

平成30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・人工知能実装研究事業))
分担研究報告書

自然言語処理を含む機械学習に供するための標準データを電子カルテから抽出するための研究

研究分担者 興相 貴英 自治医科大学・医療情報部・教授

研究要旨

自然言語処理など計算機を用いて複数施設から得られた電子カルテ情報を分析するためには予め各施設で標準的なフォーマットで出力する必要がある。しかし、電子カルテベンダー間でデータベースは全て互換性がなく、また実際に処理対象としたいデータは電子カルテデータベースではなく、周囲の部門システムに含まれていることが多い。そのため、本研究では共通フォーマットである SS-MIX2 形式として、電子カルテからは検査・処方情報を、部門システムからは循環器関連情報を抽出し・格納し、なおかつ必要な条件で SS-MIX2 ストレージからデータを抽出するシステムの構築及びデータ抽出を行い、評価した。

A．研究目的

近年大規模病院では電子カルテの導入が進み、医療データは最初から電子データとして格納されていることが多い。実際に、単施設内でデータ抽出を行い、二次利用を行っている事例は少なくない。しかし、自然言語処理など計算機を用いて複数施設から得られた電子カルテ情報を分析しようとした場合、施設間でデータ形式が異なることから、統合して用いることは容易ではない。そのため、本研究では元々地域医療連携のためのフォーマットとして考案された SS-MIX をさらに拡張した SS-MIX2 形式で電子カルテおよび複数の循環器モダリティからデータを出力・格納し、さらに格納ストレージから必要条件でデータを検索・抽出するシステムの構築を目的とした。

B．研究方法

当施設では既に電子カルテデータを SS-MIX2 に出力する機能は構築しており、本研究ではさらに心電図データ、心エコーデータ、心臓カテテル検査レポートデータについて SS-MIX2 に出力するためのシステムを構築する必要があった。心電図データ、心エコーデータ、心臓カテテル検査レポートデータについては施設間をまたいだ共通フォーマットとして日本循環器学会と保健医療福祉情報システム工業会が共同で策定している SEAMAT に準拠したフォーマットを採用した。心電図については心電図サーバから SS-MIX2 拡張ストレージに SEAMAT 形式で出力する機能を付加した。心エコーデータについては、レポートサーバに

SEAMAT 形式で出力する機能を実装することが現時点で困難であるため、データを CSV として出力した後、日本循環器学会で開発された変換プログラムを用いて SEAMAT 形式に変換した後に SS-MIX2 拡張ストレージに出力した。心臓カテテル検査レポートデータについては本研究に参加する5施設が共通で用いている CAIRS フォーマットで出力した後、SEAMAT 形式に変換した。

虚血性心疾患の危険因子についてはカルテ記載部分に含まれており、SS-MIX2 標準ストレージには出力されないため、電子カルテデータベースから SQL を用いて抽出した。危険因子に関する記載はほとんどがテンプレートにより入力されていたため、プログラムの処理してデータを整理して取り込むことができた。

イベントデータについては教師データも兼ねて人手でカルテを確認してデータを取得・入力した。

さらに SS-MIX2 ストレージから条件を指定して抽出できる SS-MIX2 agent システムの構築を行い、異なる施設でも同様の手順でデータを抽出できるようにした。

本研究では既存情報のみを利用したため、倫理委員会の承認を得た上で、病院及び研究室のホームページに研究内容、対象者を記した通知文を掲示し、対象者がオプトアウトできるようにした。

(倫理面への配慮)

人を対象とする医学系研究に関する倫理指針を遵守し研究を遂行する。

C．研究結果

対象疾患として主に虚血性心疾患を選択したため、心臓カテーテル検査を受けたことがある患者約 3000 名をデータ抽出対象とした。電子カルテから抽出した処方データについては当初想定した通りのデータが抽出できた。血液検査値データについては LDL-C 等のデータに一部欠測が認められたため、SS-MIX2 抽出システムを見直し、改めてデータを抽出した。

心エコーデータについては CSV データから SEAMAT 形式に変換して SS-MIX2 拡張ストレージに出力できた。心臓カテーテル検査レポートデータについても CAIRS-DB から CAIRS フォーマットで出力したデータを SEAMAT 形式に変換できた。カルテテキストデータについても電子カルテデータベースから抽出できた。

さらに SS-MIX2 ストレージに格納された各種データを SS-MIX2 agent を用いて抽出することにも成功した。

D．考察

今後さらに多くの施設でこうした共通フォーマットによるデータ抽出が必要とされるが、現時点では心電図サーバや心エコーレポートサーバから直接標準フォーマットに SS-MIX2 に出力できる施設は多くない。今後は各施設でシステム更新時にこうした機能を実装するよう普及を図る必要がある。また、自然文で記載されたデータも重要な情報を含んでおり、自然言語処理による情報抽出も重要であるが、一方で入力時にテンプレートを用いることにより、二次利用時により正確に情報抽出できる仕組みの導入も重要かも知れない。

E．結論

複数施設で主として循環器関連の共通形式のデータ抽出ができ、自然言語処理によるイベントデータの自動抽出のための素地を整えることができた。

G．研究発表

- 1．論文発表 特になし
- 2．学会発表 特になし

H．知的財産権の出願・登録状況

- 1．特許取得 特になし
- 2．実用新案登録 特になし
- 3．その他 特になし