

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

人口知能(AI)を活用した循環器疾患の登録システムの整備に関する研究

研究分担者 西村 邦宏 国立循環器病研究センター・予防医学・疫学情報部・部長
竹上 未紗 国立循環器病研究センター・予防医学・疫学情報部・室長

研究要旨

これまでの研究において、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、カルテ記載内における「症状記載」について、判別・予測する方法論の検討を行った。臨床研究業務担当者が実際に必要とする症状記載データについて、カルテ記載情報から手動で抽出を行った。これらを用いて教師データを作成し、カルテ記載における「症状記載」と「その他の記載」についての自動判別器を作成した。こちらは感度・特異度はともに70~80%の性能を有していることがわかった。

この結果をもとにさらに自然言語処理技術に関して先進的なIBMワトソンによりMajor Cardiac eventをとらえることを目的に辞書チューニングを行った。心筋梗塞レジストリMIDAS研究を中心とした約2000人の国立循環器病センター入院患者に関して、最も記述が的確と考えられる退院時サマリの記述をもとに虚血性心疾患、心不全、脳卒中、心臓死、全死亡に関してIBMワトソンエクスプローラーにより抽出を行った。死亡イベントに関しては、電子カルテ上の死亡退院により100%の把握が可能であった。初回の入院に関しては、入院契機が虚血性心疾患、心不全、脳卒中である場合もほぼ捕捉可能であった。死亡と入院契機の虚血性心疾患、心不全、不整脈項目により心臓死の確認が可能であった。辞書チューニング前はaccuracyとして65%前後であるが、チューニング後は95%以上の精度達成が可能であった。最終的に、ピナクルレジストリの項目中、完全自由記載の5%を除く95%の項目を抽出することに成功した。

入院時における血性心疾患、心不全、脳卒中、心臓死、全死亡に関してIBMワトソンエクスプローラーによる自然言語処理でほぼ正確にMACEの捕捉が可能であった。

A. 研究目的

詳細な臨床情報が研究、臨床試験への活用には多数(数十-数百項目)臨床情報の登録が必要であるが、しかし多忙な医療関係者の手入力は、負担が大きく大規模の臨床レジストリーの登録は困難である。一方医師に限らず、手入力による情報の入力欠損値、誤入力の増加、さらには登録数自体の低下につながることも多い。そこで近年発達してきた人工知能による自動入力システムの構築を試みた。

対象としては前項と同様に、既存の電子カルテ(主にサマリー情報)から循環器疾患レジストリー構築に必要な情報をIBM社ワトソンの人工知能により、自動抽出するシステムを構築し、特に電子カルテ情報を用いて、Major Adverse Cardiac Event(MACE)などのイベントを発生時期とともに抽出することを主目的としたシステム構築を行った。

B. 研究方法

IBMワトソンエクスプローラーを用いた抽出に関しては下記の例のように、院内症例に関して辞書定義データウェアハウス内で匿名化された電子カルテデータからのテキストマイニングによる情報抽出辞書チューニングを繰り返して精度を高めた抽出ロジックを構築した。(図2)この際用いたのは自然言語処理(natural language processing: NLP)と呼ばれる技術であり、人間がふだん日常的に使っている自然言語をコンピュータに理解させる技術であり、アメリカ癌学会など多くの欧米の学会で自動入力システムの開発が行われている。

こちらについても、MIDAS研究登録患者の電子カルテ記事から人工知能応用による自然言語処理を用い、胸痛、浮腫など症状の有無(大項目8項目、小項目25項目)について自動抽出が可能か検討した上で予測因子としての有

用性を検討した。専門医 2 名、統計学者 2 名が辞書チューニングを行い、AI による教師あり学習をサポートした。入院中に自由記載されたカルテの SOAP 記事を使用した（看護記録含む）約 60 万行のカルテ記事を読み込み、症候の出現頻度を患者ごとに集積した。

さらに AHA/ACC の PINNACLE レジストリーが循環器の疾患登録に必要な基本情報をほぼつくっていることから自動抽出により、約 216 大項目のうち何%が抽出可能かを検討した。

（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

先行する安田班の研究では検査値、患者基本情報などから抽出可能な因子数は 216 項目中 117 項目（54.2%）であった。今回自然言語処理による項目抽出をこころみたと、専門医に辞書チューニング後はほぼ医療職から見て問題ない精度の症候抽出は可能であった。これまで約 2 年間の継続した辞書チューニングにより、約 95%までの項目の抽出が可能になった（図 1）

さらに拡張要件として、冠動脈イベント等の日付けデータを自動抽出するロジックを開発した。これを用い、患者ごとに心血管系の ACS、心不全入院等のイベント自動抽出し、時系列に表示させるプログラムを試作した。こちらは対象イベントごとに根拠となる文例も同時に確認できるため今後臨床研究を行う際のデータマネジメントに関する作業効率の大幅な向上につながる。

さらに現在 ACS 6 千例、HF2000 例程度の既存の登録済みデータと照合し validation の作業中であり。本年度 6 月に論文投稿の予定である。

D. 考察

ワトソンなどの自然言語処理による自動入力システムの構築は、現在高騰しつつある臨床試験などの登録コストの引き下げにつながる可能性がある。疾患レジストリーの構築には通常数千万から臨床治験など数億円が投入、データ管理に必要となりつつあり、基本的な臨床情報を抽出し、さらに臨床試験への対象に合致するか否かの case finding などにも応用が可能な技術と考えられ、特定臨床研究等の推進にも寄与できると考えられる。

E. 結論

入院時における血性心疾患、心不全、脳卒中、心臓死、全死亡に関して IBM ワトソンエクスプローラーによる自然言語処理でほぼ正確に MACE の捕捉が可能であった。

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 |

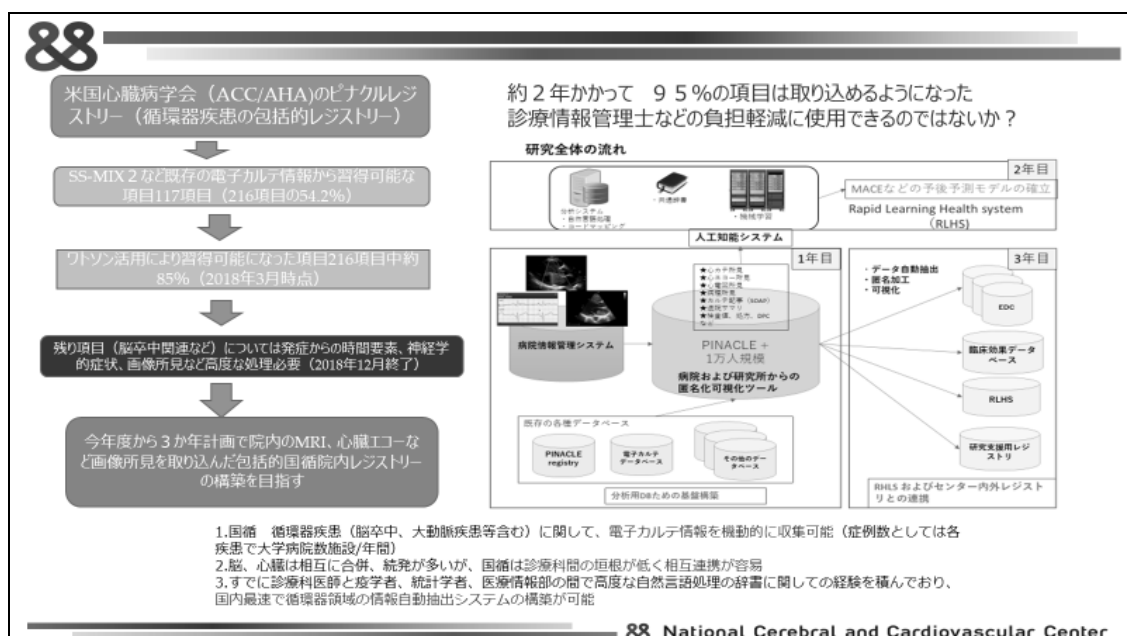


図 1 人工知能による米国心臓学会ピナクルレジストリーの項目の自動入力システム

