

共同企画

共同企画10

日本循環器学会：循環器領域におけるビッグデータ活用

2018年11月25日(日) 09:00 ~ 10:30 E会場 (5F 501)

[4-E-1-1] IBM ワトソンを利用した電子カルテからの自然言語処理の試み

○西村 邦宏（国立循環器病センター）

これまで、国立循環器病研究センターの循環器疾患患者に関して、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ている。抽出には IBM[社のワトソンエクスプローラーを用いた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2000名、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ところ、ほぼ医学的に問題ない精度の症候抽出は可能であった。さらに抽出した症候を既存の予測モデルに加えた場合10%程度の予測精度の向上を認めた。現在までに米国AHA/ACCによる循環器疾患に関する総合的なレジストリーであるNCDRのPINNACLE registryについて約400項目のうち約55%は既存の構造化データから自動抽出が確認であり、既往歴、身体所見、症状、など更に本年度の8月時点で約100項目の情報を電子カルテの情報から抽出し、総項目の約85%の抽出が可能となっている。記述の形態素解析、構文分析（特に日本語における係り受け）などの強みを持つ点が技術的特色であり、表記ゆれを含めた多義性に対して専門医による辞書チューニング作業を加えて、精度向上を行っている。今後2年程度で脳卒中を含めた循環器領域に関しては、心カテ、心エコー所見などの電子カルテ部門システムにわたる詳細項目を含め、臨床研究および将来的な臨床試験への応用を見据えた自動抽出システムを構築可能である。さらに抽出した情報をもとにMACE(Major Adverse Cardiac Event)の発症を人工知能により予測可能かを検討している。本年度からは冠動脈不安定プラークの心臓MRI画像からの自動抽出アルゴリズム開発など画像解析にも応用している。

循環器領域におけるビッグデータ活用

中山雅晴^{*1}、西村邦宏^{*2}、
中井陸運^{*3}、武田理宏^{*4}、的場哲哉^{*5}、

- *1 東北大学大学院医学系研究科医学情報学、*2 国立循環器病研究センター研究開発基盤センター、
*3 国立循環器病研究センター循環器病統合情報センター、*4 大阪大学医学部附属病院医療情報部、
*5 九州大学病院循環器内科

Leverage of Big Data in Cardiology

Masaharu Nakayama ^{*1}, Kunihiro Nishimura ^{*2},
Michikazu Nakai ^{*3}, Toshihiro Takeda ^{*4}, Tetsuya Matoba ^{*5}

*1 Medical Informatics, Tohoku University Graduate School of Medicine,

*2 *3 National Cerebral and Cardiovascular Center, *4 Dept. of Medical Informatics in Osaka University Hospital

*5 Angiocardiology Kyushu University Hospital

Utilization of medical big data is expected to bring about science and technology innovation. However, the data available for integration and further analysis are limited. In a joint session of the Japan Cardiology Society and Japan Medical Informatics Association, we will discuss the current availability of big data in cardiology and the limitations to use. Dr. Mishimura will discuss natural language processing of electronic medical records using Watson. Dr. Nakai will discuss a project using DPC Data managed by the Japan Cardiology Society (JCS) and the Japanese Registry of All Cardiac and Vascular Diseases. Dr. Takeda will present a case reporting system linked to clinical information, image data, and bio-samples. Dr. Matoba will report on J-IMPACT, which is developing a network system to collect clinical data in cardiology, including catheter examinations and outcome events. Lastly, Dr Nakayama will present a standard format authorized by JCS to transfer numeric electrocardiogram and echocardiogram data from a physiologic examination system to Structured Medical Information eXchange version 2 extended storage.

Keywords: Watson, JROAD, CDSC, J-IMPACT, SEAMAT

1. 緒論

DPC やレセプト、電子カルテ等からのデータ収集が進むにつれ、医療ビッグデータとしての活用が期待されている。しかしながら、臨床医としては各専門分野におけるデータの内容とクオリティが担保されていることが重要である。本シンポジウムでは、循環器疾患分野において、大規模にデータを収集もしくは電子カルテからの情報を有効活用に取り組む研究者を集め、現状を報告する。具体的には、人工知能を活用したデータ解析、全国循環器専門施設からのDPCとレセプトのデータを活用した JROAD、カテーテルデータを中心としたデータを集める J-IMPACT 事業、研究における画像収集システム、心電図や心臓超音波検査のデータを活用するための循環器学会標準出力フォーマット(Standard Export data format: SEAMAT)などのテーマで循環器学会が直面する課題や今後の発展を提示していただく。本セッションを通して、循環器学会と医療情報学会および関連組織の進むべき方向性と進捗を確認し、広く意見交換を行いたい。以下、講演者それぞれの概要を記載する。

2. IBM ワトソンを利用した電子カルテからの自然言語処理の試み (西村)

これまで、国立循環器病研究センターの循環器疾患患者に関して、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ている。抽出に

は IBM[社のワトソンエクスプローラーを用いた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 2000 名、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ところ、ほぼ医学的に問題ない精度の症候抽出は可能であった。さらに抽出した症候を既存の予測モデルに加えた場合 10%程度の予測精度の向上を認めた。現在までに米国 AHA/ACC による循環器疾患に関する総合的なレジストリーである NCDR の PINNACLE registry について約 400 項目のうち約 55%は既存の構造化データから自動抽出が確認であり、既往歴、身体所見、症状、など更に本年度の 8 月時点で約 100 項目の情報を電子カルテの情報から抽出し、総項目の約 85%の抽出が可能となっている。記述の形態素解析、構文分析(特に日本語における係り受け)などの強みを持つ点が技術的特色であり、表記ゆれを含めた多義性に対して専門医による辞書チューニング作業を加えて、精度向上を行っている。今後 2 年程度で脳卒中を含めた循環器領域に関しては、心カテ、心エコー所見などの電子カルテ部門システムにわたる詳細項目を含め、臨床研究および将来的な臨床試験への応用を見据えた自動抽出システムを構築可能である。さらに抽出した情報をもとに MACE(Major Adverse Cardiac Event)の発症を人工知能により予測可能かを検討している。本年度からは冠動脈不安定ブランクの心臓 MRI 画像からの自動抽出アルゴリズム開発など画像解析にも応用している。

3. 循環器疾患診療実態調査(JROAD)の現状と進捗 (中井)

循環器疾患診療実態調査(The Japanese Registry Of All cardiac and vascular Disease: JROAD)は、日本循環器学会の主導により循環器疾患診療に関する我が国の現状を把握するため、2004年より循環器科・心臓血管外科を標榜する施設を対象に実施している調査である。調査内容は、施設ごとの病床数・医師数など施設規模情報・急性心筋梗塞や心不全など循環器疾患患者数・検査件数・治療件数データの収集を行っている。2012年調査よりデータセンターを国立循環器病研究センターに移管、運用を開始し、循環器研修施設・研修関連施設より100%の登録率を達成している。

また、2014年度よりDPC(診断群分類包括評価)対象施設より循環器疾患患者の個別症例情報の収集を目的としたDPC情報として登録を開始している(JROAD-DPC)。収集している項目として、性別・入院時年齢・入院時診断名・入院時併存症名・入院後合併症名とそれらのICD10コード・手術処置名とその実施日・使用された薬剤・在日数や医療費などの情報である。現在で2012年度-2016年度(4年分)の3,626,656症例のレコード数が収集されており、急性心筋梗塞症の年間入院患者数が約7万人で、その入院死亡が8%である。さらに、心不全の入院患者数は年間20万人を超え、年度毎に増え続けている。JROAD-DPCは、同一病院に再入院・通院している症例でないというDPCの特性の1つがあるものの循環器疾患における全国規模のビッグデータであるといえる。それらの蓄積されたデータを循環器疾患の医療の向上につなげるために、日本循環器学会は学会員から研究テーマの公募を行い、公募3年目の現在までに43課題(初年度=11課題、次年度=16課題、今年度=16課題)を採択した。研究進捗としては、多くの論文課題が論文作成の段階まで到達している。今回は、JROADの概要とデータの構造とともに、JROAD研究の申請手順と現状の研究進捗状況について発表する予定である。

4. 電子症例報告システムと連携する臨床画像収集システム、臨床サンプル管理システムの構築 (武田)

我々は、電子カルテとデータ連携したCDISC標準準拠の電子症例報告システム(CDCS)を構築し、OCR-net(大阪臨床研究ネットワーク)で連携する15の医療機関に導入している。今回、CDCSと連携する臨床画像収集システム、臨床サンプル管理システムを開発した。

臨床画像収集システムはDICOM画像については、CDCSで指定された画像撮影日、モダリティから対象画像を検索し、患者IDをCDCSで付番される被検者IDで匿名化して、データセンターに送付する。非DICOM画像については、特定のフォルダに画像を保存することで、ファイル名に被検者番号が付与した形でデータセンターに画像を転送する。この際、あらかじめ個人情報が含まれる領域を指定すれば、個人情報が含まれる流域をマスクする機能を有している。

臨床サンプル管理システムについては、プロジェクトコード、施設コード、プロジェクト内患者番号、サンプル種番号、材料種別番号、枝番で定義される臨床サンプルIDを発番すると共に、患者IDから研究用患者IDを発番する。データセンターでは、臨床サンプルIDとCDCSで発番される被検者番号、研究用患者IDの対応表を保持する。研究用患者IDは別プロジェクトで取得されたサンプル情報を紐づけるために利用さ

れる。センターサーバでは臨床サンプルIDを遺伝子サンプルIDに二重匿名化することが可能である。臨床研究者、遺伝子研究者は本システムを用いてそれぞれのサンプルIDで互いに問い合わせをかけると、システム内で臨床サンプルIDと遺伝子サンプルIDが変換され、相手方に問い合わせをすることが可能となる。

5. 心臓カテーテルを中心とした多モダリティ循環器診療情報を収集するJ-IMPACTシステム (的場)

循環器疾患は心筋梗塞、脳卒中を代表とした血管病により日本人の生活の質を損ねる重大な疾患領域である。特徴的に発展した心臓カテーテル検査・心臓カテーテルインターベンション治療の診療情報、また、大規模臨床試験のエビデンスに基づく薬物療法などの診療情報などは、構造化データとして個別患者のリスク予測や治療プロセスの再評価に利用されるべきであるが、多モダリティの診療情報を電子的に収集するシステムは存在しなかった。我々は日本循環器学会およびImPACT事業の支援を受け、電子カルテにおける患者基本情報、処方、検体検査データをSS-MIX2標準ストレージから、また、生理検査や心臓カテーテル検査・心臓カテーテルインターベンション治療レポートの情報をSS-MIX2拡張ストレージから収集するJapan Ischemic Heart Disease Multimodal Prospective Data Acquisition for Precision Treatment (J-IMPACT)システムを開発してきた。本発表ではJ-IMPACTシステムの現状と課題を議論したい。

6. SEAMATでできること、導入のためにすべきこと (中山)

Structured Medical Information eXchange version 2 (SS-MIX2)を活用したデータベース及びそれを利用した研究が広まってきているが、用いられるデータはSS-MIX2標準化ストレージ由来が一般的で、患者基本情報、病名、処方、採血検査結果が代表的項目である。しかしながら、臨床においては診療科特有のデータが診断や治療に重要であり、その活用が望まれる。そのためIHE-J循環器や日本循環器学会IT/DB委員会を中心に、心電図や心臓超音波検査、カテーテル検査・治療のデータをSS-MIX2拡張ストレージに出力するための標準フォーマットであるStandard Export data format (SEAMAT)を定めた。検査毎に標準的な項目名や単位、およびLOINCコードが当てられており、その情報は日本循環器学会ホームページに掲載されている。項目策定にあたっては、日本循環器学会のみならず、心不全学会、不整脈心電学会、心エコー図学会、心血管インターベンション治療学会、心臓核医学会、医療情報学会が協力した。既に、複数のベンダーがSEAMATに従った製品を販売しているため、施設にSS-MIX2ストレージが確保されていれば、生理検査システムやカテーテルレポートを新規に購入する際に仕様を準拠させることで導入が可能である。さらに、我々はデータ自体がcsvやxmlファイルであればSEAMAT形式へ変換するためのプログラムも作成した。また、上記各学会からのメンバーが集まり、SEAMAT研究会として見直しの体制も確立している。例えば、心電図に関しては、安静時12誘導だけとなっているので、ホルター心電図検査や負荷心電図の項目策定を検討し始めている。SEAMATにより目指すものは、循環器特有検査におけるデジタル化されたデータの転送および活用であり、循環器臨床医のニーズを満たすものと考えている。