共同企画

共同企画4

日本循環器学会：循環器領域におけるビッグデータ活用の現在

2017年11月22日(水) 13:30 ~ 15:30 F会場 (10F 会議室1004-1005)

[3-F-2-JS4-1] 人工知能応用による自然言語処理を活用した電子カルテのセマンティック登録と 全国登録事業への将来展望

西村 邦宏（National Cerebral and Cardiovascular Center）

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た。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75%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。

さらに、1年後の 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た。古典的サポートベクターマシンおよびランダムフォレストの複数モデルを組み合わせたアンサンブル学習により、MACE予測モデルを構築した。学習用データセットでは偽陽性5%程度、偽陰性立5%程度で予測するモデルが生成可能であり、検証用データセットでも高リスク群,低リスク群に関して AUC0.88で予後予測が可能であった。この成績は GRACE研究などの既存の MACE予測モデルの成績より感度、特異度ともに約20%程度向上していた。

今回はこれらの結果を含め、セマンティックスを活用した今後の全国登録への展望について報告する

循環器領域におけるビッグデータ活用の現在

中山雅晴*1、西村邦宏*2、中村文明*2、武田理宏*3、興梶貴英*4、
松村泰志*3、小川久雄*2

*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、

*3 大阪大学大学院医学系研究科医療情報学

*4 自治医科大学医療情報学

Big Data in Cardiovascular Medicine

Masaharu Nakayama*1, Kunihiro Nishimura, Fumiaki Nakamura, Toshihiro Takeda*3,
Takahide Kohro*4, Yasushi Matsumura*3, Hisao Ogawa*2

*1 Tohoku University, *2 National Cerebral and Cardiovascular Center,

*3 Osaka University, *4 Jichi Medical University.

In the era of big data, the utilization and analysis of significant amounts of clinical data are imperative. Here we have a symposium to show the current status of utilizing a variety of clinical data for cardiovascular research in co-operation with The Japanese Circulation Society and Japan Association for Medical Informatics. Five speakers will present each project. Dr. Nakamura will provide information on a nationwide database using the administrative case-mix Diagnostic Procedure Combination (DPC) system (i.e., the Japanese Registry of All cardiac and vascular Diseases (JROAD)-DPC) to reveal the current status of cardiovascular medicine in Japan. Dr. Nishimura will present his research using natural language processing with artificial intelligence for semantic registration from electronic medical records. Dr. Takeda will talk about a case reporting system that enables electronic data capture from electronic medical records according to the CDISC format. Dr. Kohro will talk about J-Impact, in which tons of data from health insurance claims and electronic medical records including catheterization and intervention data are utilized to reduce the patient risk of cardiovascular diseases. Dr. Nakayama will describe a standard export data format (SEAMAT) authorized by the Japanese Circulation Society. SEAMAT is a format for ECG, echocardiograms, and catheterization data to be transferred to the SS-MIX2 extended storage.

Keywords: Big Data, Cardiology,

1. 概説

超高齢社会において、循環器疾患の患者数は益々増大することが懸念される。その対策として、従来からの大規模臨床試験に加え、最近ではDPCやレセプト、電子カルテデータなどのリアルワールドデータを用いたビッグデータの活用が期待されている。また、昨今の人工知能・機械学習の進歩により新規ICT技術を用いた知見創出も現実となってきた。今回、日本循環器学会と医療情報学会との合同シンポジウムにおいて、大規模データ収集の現状や進行しているビッグデータプロジェクトの詳細を明らかにする。本セッションを通して、両学会および関連組織の進むべき方向性の確認し、多くの方々とのディスカッションを経て、さらなる高みを目指すための方策を探りたい。

2. 各テーマ

国立循環器病研究センターから、循環器疾患診療実態調査(The Japanese Registry Of All cardiac and vascular Disease: JROAD)について中村が説明する。JROADは、循環器疾患診療に関する我が国の現状を把握するため、日本循環器学会の主導により2004年から循環器科・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、治療件数等が含まれる。2014年度からはDPC(診断群分類包括評価)対象施設より循環器疾患患者のDPC情報の収集を開始した(JROAD-DPC)が開始された。これにより、施設情報と患者個

人情報とを組み合わせた分析が可能となった。また、こういった結果を調査協力施設へ診療の質指標(Quality Indicator)の測定値としてフィードバックすることも行っている。2016年度の調査で、JROADには1,573の施設が参加、JROAD-DPCは739施設から938,605件のデータが収集された。蓄積された循環器ビッグデータに関しては、日本循環器学会員から研究テーマの公募を行い、現在までに25課題を採択している。上記経過と今後の展開についても概説する。

さらに、国立循環器病研究センターの西村から、人工知能応用による自然言語処理を活用した電子カルテのセマンティック登録と全国登録事業への将来展望を述べる。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、循環器疾患特有の症状の有無について自動抽出が可能か検討、ほぼ医学的に問題ない精度の症候抽出が可能であった。また、抽出した症候を既存の予測モデルに加えた場合10%程度の予測精度が向上した。さらに、1年後のMajor Adverse Cardiac Eventの発症を古典的サポートベクターマシンおよびランダムフォレストの複数モデルを組み合わせたアンサンブル学習により、予測モデルを構築し、AUC 0.88で良好な予後予測が可能であった。これらに加え、セマンティックスを活用した今後の全国登録への展望なども説明する。

また、電子カルテの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の構築と循環器領域での活用について、武田から説明する。臨床研究は多施設で行うこ

とが求められるが、既存のデータ収集システム(EDC)は電子カルテ(EMR)と連結していないため、EMR から EDC への転記作業が発生するという問題点がある。たとえば、転記のために人手が必要であること、転記ミスが生じるリスク、モニターによる診療録の確認作業(SDV)が必要であること、などが挙げられる。武田らは、EMRとEDCの二重入力を排除し、臨床データを収集するシステムを開発、CDISC 標準規格に準拠したフォーマットでデータセンターとデータの受け渡しを可能とした。EMRとEDCの連携は3社の電子カルテで実現し、データ転記型(NEC)、直接記録型(NEC, IBM)、データ転送型(FUJITSU)の連携を行っている。それらを利用した3件の臨床研究についても説明し、今後の発展予定、さらに、本システムと連携した、臨床画像の自動取得、臨床サンプルのラベル発行を可能とするシステムの導入などの計画も述べる。

興梠より現在内閣府の ImPACT プログラム「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」について説明する。心臓カテーテル検査や冠動脈インターベンション等、循環器症例データを収集・解析することによって将来の疾患リスクをシミュレートし、結果をフィードバックすることにより患者の疾患リスクを低減する研究を示す。

最後に、中山からは心電図や心臓超音波検査等から得たデータを SS-MIX2 拡張ストレージにデータを出力するために定めた標準規格、日本循環器学会標準出力フォーマット(Standard Export data format:SEAMAT)についての報告を行う。厚生労働省電子的診療情報交換推進事業によって

定められた Standardized Structured Medical Information eXchange(SS-MIX)ストレージ形式および2版となる SS-MIX2 は現在普及が進み、2016年度末で全国630施設に導入されている。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採血検査値、処方などのデータが格納され、標準化コードも備わっている。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り、データの2次利用は困難な状況となっている。心不全や心筋梗塞といった疾患を扱う循環器分野においては、心電図や心臓超音波、心臓カテーテル検査といった複数のモダリティが重要である。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目的で、IHE-J循環器を中心とした有志は日本循環器学会に働きかけ、標準化へ向けた取り組みについて理事会の承認を得た。結果、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 forMAT (SEAMAT)を公式に定めた。項目の策定にあたっては関連学会の協力を得て、多くの施設で共通する項目を中心に定めた。今後各企業より SEAMAT に対応した製品が商用化され、多くの循環器系臨床データの効率的な抽出が可能となる。これにより、臨床研究においてデータ転記が参加医師や研究補助員にとって大きな負荷となっていたが、その軽減およびより大規模で正確性のあるデータ収集が期待される。

座長は日本循環器学会より国立循環器病研究センターの小川久雄先生、日本医療情報学会からは大阪大学の松村泰志先生が務める。