

◆Summary

Expectation to DWH using a novel natural language technology
Use of the non-structured text information in the healthcare setting is thought to be essential not only to confirm medical treatment information and effective use but also to support medical treatment and support writing medical treatment record and standardization. By a new natural language analysis technology, these expectations are in the field of vision.

検証

DWH導入がもたらす メリットと課題

自然言語解析技術組み込みの DWHへの具体的な期待

宇宿功市郎

熊本大学医学部附属病院 医療情報経営企画部長

要旨・医療現場での非構造化テキスト情報の活用は、診療情報の確認、効果的2次利用のみならず、診療現場での支援、診療記録の支援、標準化にも必須と考えられている。新たな自然言語解析技術により、このことの実現が視野に入ってきている。

医療現場での情報の電子化の進展には目を張るものがあり、患者の移動履歴や検査履歴、コード化された病名などは標準化された形で格納がなされ、検査結果などの数値情報、処方箋の内容と共に院内での情報共有、地域によっては地域内での情報連携が可能となってきた。場合によっては加工処理しての2次利用も検討が進んできている。

また、2次利用を考える際には、情報の取り漏れを防ぐ意味もあり、項目を選択抽出しての情報入力テンプレートの作成と入力依頼が行われてもいる。このような情報は検索を行いやすくするために、ある一定のルールに従って入力されているもので、一般には「構造化データ」と呼ばれている。

ただ、さまざまな場面で発生する情報の全てが構造化されているわけではなく、どうしても記録できない情報、記録できたとしてもすぐには解析ができない情報があるのも事実である。これまでのいくつかの検討によると蓄積された情報の7〜8割は構造化が成されておらず、解析の際の問題点となっているとの指摘がある。

構造化された診療関連情報の解析では、保険診療関連の検討はかなりのところまで行えるようになってきている。しかしながら、診療現場で発生している可能性がある事象の中で情報として捉えられていないもの、情報を得ようとするとな大な労力（仕組みの構築や現場での入力負担の増大など）を要し、なかなか実現が難しいものがあるのも事実である。準備、工夫にもかかわらず構造化ができない場合もあり得るのである。

理由は、医療現場のみならず多くのことが生じる際に発生する情報、その内容を全て予測することは不可能であり、加えて訓練した

集団からのみでの入力に限らなければ、収集した情報への信頼性が十分ではなくなる場合があるからである。

「非構造化データ」解析研究や 応用への流れ

これらを解決するために、構造化されていない情報、すなわち一般に「非構造化データ」と言われるものの中からある事柄の情報を構造化とはいかないまでも、構造化に近い形にまとめることや変換しての情報の収集や解析のための工夫が行われている。この作業はとても大変なもので、さまざまな技術が用いられ、足りない場合には新技術の開発も取り組まれ、1つの学問ともなっているとある。つまり、この非構造化情報の取り扱いだけでも十分に研究対象や業務の1つとなっているのである。

このことは素晴らしいことであると同時に、注意が必要な点でもある。診療現場の人



1985	1995	2005	2015	2025	2035
医事会計システム導入					
統合オーダリングシステム導入					
データウェアハウスDWH導入					
1999電子保存3原則/電子化診療録導入					
2005医療情報ガイドライン(現在は第4.3版2016.3)					
2005個人情報保護法					
2015介護保険法改正,地域包括ケア					
2016マイナンバー運用開始,今後の拡がり?					



図1 医療情報システムの発展—これまでと今後

間が簡単には取り組めないものになってしま
うことが少なくならずあるからである。もち
ろん技術の進歩で、初心者とはいかないま
も、ある程度の知識があれば使いこなせるも
のがここ数年現れてきており、ビッグデー
タの解析の1つとして紹介され、さまざま
な場面で活用が取り組まれている。

自然言語解析技術 「iKnow」活用の経緯

さて今回は、これまでとは少しばかり趣の
異なった自然言語解析技術である
InterSystems iKnowを病院内のデータウェア
アハウス (Data Warehouse : DWH) と同
時に活用して、何を指しているのか、何が
目指せるのかについての我々の取り組み、考
えを紹介させていただきたい。

病院情報システムは約35年程前の医事会計
システムの導入から始まり、その後検査等
の指示や結果の表示の仕組み (一般にオー
ダリングと呼ばれる) として発展を遂げ、20
00年代に入り診療における医師をはじめと
する医療者の記録が電子的に行われるよう
になり、現在の電子カルテの状況となってい
る (図1)。

この情報電子化の流れを見てお分りのよ
うに、電子化された情報は、コンピュータの
能力のこともあり、別々の仕組みにバラバラ
に蓄積される仕組みとなつて、発展してきて
いる。

余談ではあるが、05年以降は地域医療での
情報連携が模索され、近年は健康時からの情

報蓄積も行われている。このような中、今後
もテキスト情報の増加が起きることも想像が
できる。全ての情報の標準化は困難だからで
ある。また保険診療においては、月毎に集計
し会計処理を行っており、医事会計情報と検
査結果、処方情報を俯瞰した形で、さらには
患者の一診療期間で解析することができず
いた。このことを改善する目的で、1990
年代の終わり頃からDWHの導入が行われて
いる。

このようなこともあり、DWHの利用は医
療経済的面で活用、その後は疾病登録、コ
ホート研究などの臨床研究における活用も
進み発展してきている。入力情報の標準化の
流れとも相まって、発生した事象を中心に解
析を行う研究の進歩には大変有用な仕組みと
なっているのである。

ただ先にも述べたが、電子カルテ内の情報
が全て構造化されているわけではなく、何を
どのように判断したので、ある医療行為が行
われた、必要だった等の内容は今でも非構造
化のまま入力力がなされている。したがって、
行われた診療内容・行為を時系列に追跡する
ことで、診療の実情を把握、その情報を利用
しての研究が行われているのである。

テキスト情報解析の 取り組みと目的

今回はこのような検討に加えて、診療記録
に電子的に記載されているテキスト情報を、
これらの情報と組み合わせることで新たな発
見につなげていくことができないかと期待

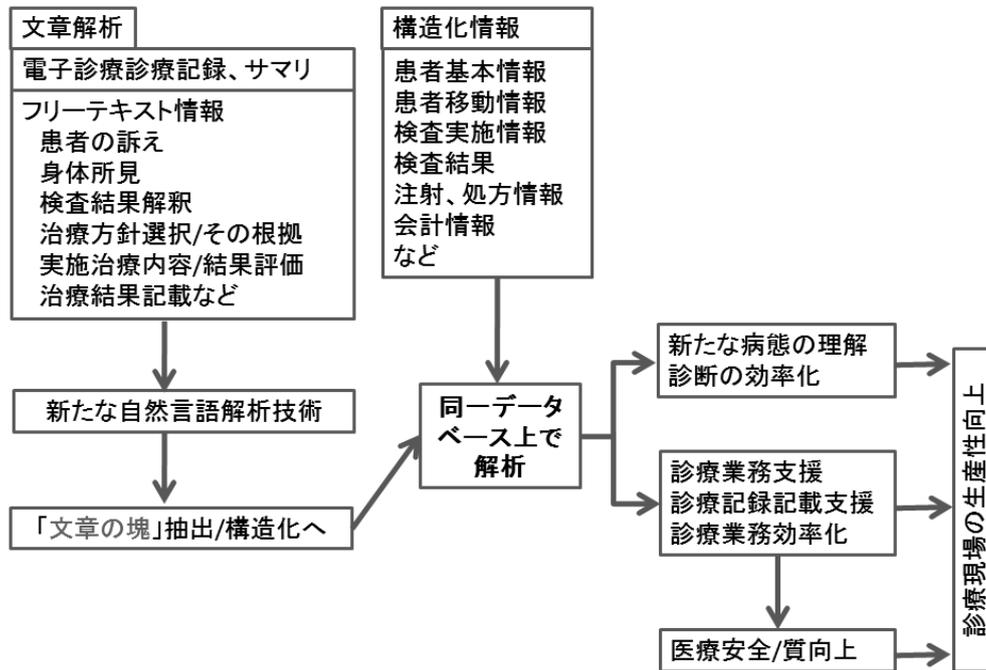


図2 テキスト情報活用による診療現場の生産性向上

し、取り組んでいるところである。診療現場で入力されるフリーテキスト情報を疾患ごと、疾患の病態出現の度に収集、自然言語処理を行い、電子化診療録に記載のある語句の中から症例、病態ごとに語句と意味・内容を保持した「意味のある集まり」を抽出できないかと考えているのである。

今回活用する技術は、人工知能の自然言語処理・機械学習技術を使った標準化の技術とされており、これまでの分かち書きの形態素解析とは異なったものである。事前に解析用の辞書などを準備する必要がないものであり、医療現場での利用には適していると考えられている。

従来の形態素解析においては、単語・文字として分かち書きが行われ、1つの単語の存在は確認することができても、その語句がどのような文脈で使用されているのかは不明であった。例えば、「副作用での肝機能悪化が改善した」という短文では、「副/作用/での/肝機能/悪化/が/改善/した。」となり、「肝機能」という単語の出現や頻度は確認できても、どのような文脈で単語が使われたかは解析ができなかった。どのような意味で使われているかなどの解析のためにはあらかじめ辞書等の設定が必要であり、解析結果が事前に準備する辞書の内容による設定や、その精度に影響を受けるものであった。

今回の技術は、前後の文脈から内容に基づいてテキスト分析を行い、「副作用/での/肝機能悪化が改善した。」という単位での分析が可能となるもので、フリーテキストの検討に適していると考えている。このような意

味のある塊としてのテキスト情報に、これまでに構造化されている検査結果、医療行為の情報を結びつけることが可能となれば、より詳細な病態の解析を可能とするばかりではなく、その病態を医療者がどのようにとらえ、何を考え、その行為を行ったのかの解析を可能とできるのではないかと期待しているのである。

より具体的に述べるならば、ある患者において何らかの訴えがあり、その際に症状・症候がとらえられて、診断に至るもしくは必要な検査を加え確定診断になり、何らかの思考に伴って治療を選択することになるのであるが、この流れをより可視化することに活用したいのである(図2)。

研究や診療において「KNOW」に期待するもの

「Know」は、電子化診療録のフリーテキストの中から目的とする概念を効率的に抽出することが可能で、臨床研究・治験などの対象コホートの選択を容易にできると考えている。また目的とする概念、特に病態を容易に抽出できることは、患者愁訴、身体所見、検査・治療の記述を活用しての傷病名の標準化を行え、この過程で頻度が高く記載されている意味のある文言、文節、短文をとらえることができる、特定の病態において記載が必要となる内容が明らかとなっていく可能性が出てくるのである。逆に、特定の病態において、患者の状態に合わせて確認すべきことを提示することができる可能性があり、また必

要な際には記載の支援を行うこともできるかもしれないのである。日々診療で行っていることを自動で整理、患者状況に合わせて、医療者の思考支援、実施記録支援を行う仕組みとして発展させていくことを期待しているところなのである。

以上の過程の中では、機械学習を組み合わせ、カテゴリ化や重み付けを行いつつ、医療用語の標準化や、コーディングにつなげることも不可能ではないと考えている。

テキスト情報活用による 診療現場の生産性向上

医療の現場のさまざまな場面における業務では、必要かつ十分な情報を素早く得て活動を行い、そして活動内容を過不足なく記録することが求められている。院内情報の電子化においてこれまでに実現できていることは、

情報を得るための手段の開発と2次利用のための情報の標準化である。

これで、診療行為のかなりのところまでは包含できるようになっているのは事実である。ただ、これまでに標準化ができていない、もしくは行われてきていなかった情報があることも現実である。現場では、記録を残すための教育、訓練がなされてきているが、十分かと言えばまだまだ不足している面も生じている。加えて、記録を残すため、もしくは記録の中から必要な書類を作成するために費やす時間はかなりのものとなっている。

1枚の診療情報提供書、1枚の保険関係の書類作成も、その作成にはある程度の時間を要する。これが1日に十数枚ともなれば、それだけでかなりの時間となり、これが勤務医をはじめとする医師、医療者の実働の時間を延ばしているのである。

電子化された診療記録の中で、標準化の済

んでいるもの、テキスト情報で非構造化の情報を含めた文節の塊として整理、標準化、さらにはコード化まで行うことで、診療経過での情報収集提供の支援のみならず、診療実施後の記録作成の支援も行うことができるようになる。医療現場での記録の精度向上に加え、1人の医療者が医療行為を行う余地を増やすことも不可能ではなく、生産性の向上につながるものと期待している。

※ ※

宇宿功市郎（うすく・こういちろう）●57年熊本県生まれ。81年鹿児島大医卒。同大医学部第3内科、同大大学院医歯学総合研究科准教授（人間環境学講座医療システム情報学）などを経て、06年より熊本大教授。同大医学部附属病院医療情報経営企画部長、総合臨床研究部研究データ管理センター長、大学院医学教育部医療情報医学分野を担当。