

ポスター

ポスター5

DWH

2017年11月22日(水) 08:45 ～ 09:45 L会場（ポスター会場1）（12F ホワイエ）

[3-L-1-PP5-3] 大学病院における患者情報2次活用における実情と今後の展開 ー長崎大学病院 DWHを対象とした取組ー

本多 正幸¹, 松本 武浩¹, 牛嶋 拓也², 岩元 泉², 本田 千春² (1.長崎大学大学院医歯薬学総合研究科, 2.長崎大学病院)

長崎大学病院では平成27年1月にシステム更新を行い、患者情報の2次活用システム（DWHシステム）を新規に構築し、特に構造化データ（病名、検査、処方等）の分析を可能としたシステムを構築した。構築に当たり電子カルテからのデータ抽出と取り込みに際し、負荷軽減と精度向上を目指し中間データベースと抽出ツールを用いた。また、その他の病理診断レポートや退院時サマリーシステムなどのサブシステムからの取り込みは個別のインターフェースプログラムで対応した。平成29年5月時点で実質稼働2年となっているが、患者情報の2次活用の実績としては、直近では月平均50～60件程度の実績があり、その中で病院管理統計的な帳票の依頼に対しては医事会計システム等から、また検査データを含めた患者情報の抽出・統計業務依頼は DWHから行っている。

DWHを対象とした支援内容としては、例えば、特定の薬剤を投与した患者の検査結果情報や手術（処置）情報、あるいは文書情報の特定項目に関する統計情報など、薬剤情報、検査情報、病名情報を中心に文書情報や看護必要度情報などの検索要望がある。DWHシステム導入により、便利になった部分もある反面、機能的面でも多くの要望が存在している状況である。特に、電子カルテベンダーのシステムやサブシステムのデータの粒度や範囲などに依存する、データの質の問題は今後大きな課題である。

本報告では、医療情報部の仕事の中で大きな比重と意義を占めている、患者情報の2次活用に関する業務（研究支援、業務支援）の実情を報告し、特に DWHシステムにおける機能について実績とともにその機能分析と評価を行う。また、今後の医療情報分野の大きな課題の一つとして、非構造化データ（自然言語データ）からの知見や各種アウトカムの抽出があるが、退院時サマリーの文書データに対する取り組み事例についても取り上げる。

大学病院における患者情報2次活用における実情と今後の展開

－ 長崎大学病院 DWH を対象とした取組 －

本多正幸^{*1}、松本武浩^{*1}、牛嶋拓也^{*2}、本田千春^{*2}

*1 長崎大学大学院医歯薬学総合研究科、*2 長崎大学病院・経営企画課

Current circumstances and future developments in secondary use of patient information at university hospital - Efforts targeted at DWH system in Nagasaki University Hospital -

Masayuki Honda^{*1}, Takehiro Matsumoto^{*1}, Takuya Ushijima^{*2}, Chiharu Honda^{*2}

*1 Nagasaki University Graduate School of Biomedical Sciences, *2 Nagasaki University Hospital

At Nagasaki University Hospital, system update was done in January 2015, a new secondary use system (DWH system) of patient information was constructed for especially analysis of structured data (disease name, lab-test, prescription, etc.). When constructing and extracting data from electronic medical charts, we used intermediate database and extraction tool to reduce load and improve accuracy. In addition, importing from other subsystems such as pathological diagnosis report and discharge summary system was handled by individual interface program. At May 2017, it is two years of actual operation since being activated, and as a result of secondary utilization of patient information, there is a record of about 50 to 60 reports per month on average. Among them for the request of hospital management statistical form it is derived from the medical accounting system. For the request of the statistical work of patient information including lab-test data, it is derived from DWH system.

As the support activities of using contents in DWH system, for example, lab-test result and surgery (treatment) information of a patient who administered a specific medicine and statistical information on specific items of document information. There is a search request such as document information and nursing necessity information center with usual request for drug information, lab-test information, disease name information.

Although there are some parts that are convenient due to the introduction of the DWH system, on the other hand, there are many requests on the functional side. In particular, the problem of data quality, which depends on the granularity and scope of the data of the electronic medical record system vendor and subsystem vendors, is a big problem in the future.

In this report, we report on the actual situation of work related to secondary utilization of patient information (support for research and business), which occupies a large proportion and significance in the job of the medical information department. In particular with regards to the results on functions in the DWH system, We will discuss about performance evaluation and functions in the DWH system. In addition, as one of the major problems in the medical information field in the future, there are findings from unstructured data (natural language data) and extraction of various outcomes. We refer to case examples of efforts on document data at the discharge summary.

Keywords: Secondary usage, DWH, HIS

1. はじめに

長崎大学病院では平成27年1月に、システムレスポンスの恒常的確保などの目標を掲げシステム更新を行った。以前の本学会でその概要を報告した。簡単にその概要についてまず紹介する。長崎大学病院総合病院情報システムは、2008年6月の新病棟開院に合わせて更新を行い、医科・歯科双方における電子カルテシステムや、地域医療ネットワークシステム（「あじさいネットワーク」への参入）を開始した。その中で、処方・注射システムのカレンダー化、電子クリニカルパスシステムの普及、電子指示システムの導入など HIS における先進システムへの取り組みを行ってきた。2015年1月には、基盤重点項目として、(1)システムレスポンスの恒常的確保、(2)年間を通じてのノンストップサービスの提供、(3)電子カルテ運用における3原則の担保、(4)ユーザ認証機能を含めた情報セキュリティ機能の向上、(5)患者情報の2次活用、(6)災害時復旧対策（BCP 対策）、(7)地域医療連携における中核機関としての役割、を掲げて更新運用を行ってきている。それらの目標実現のために、基盤システムとして、仮想化サーバシステム、仮想クライアントシステム、指静脈認証システム

等のインフラを導入した。

(5)の患者情報の2次活用の課題では、DWHシステムを新規に構築し、特に構造化データ（病名、検査、処方等）の分析を可能としたシステムを構築した。構築に当たり電子カルテからのデータ抽出と取り込みに際し、負荷軽減と精度向上を目指し中間データベースと抽出ツールを用いた。また、その他の病理診断レポートや退院時サマリーシステムなどのサブシステムからの取り込みは個別のインターフェースプログラムで対応した。平成29年5月時点で実質稼働2年となっているが、患者情報の2次活用の実績としては、直近では月平均50～60件程度の実績があり、その中で病院管理統計的な帳票の依頼に対しては医事会計システム等から、また検査データを含めた患者情報の抽出・統計業務依頼はDWHから行っている。

DWHを対象とした支援内容としては、例えば、特定の薬剤を投与した患者の検査結果情報や手術（処置）情報、あるいは文書情報の特定項目に関する統計情報など、薬剤情報、検査情報、病名情報を中心に文書情報や看護必要度情報などの検索要望がある。DWHシステム導入により、便利になった部分もある反面、機能的面でも多くの要望が存在して

いる状況である。特に、電子カルテベンダーのシステムやサブシステムのデータの粒度や範囲などに依存する、データの質の問題は今後大きな課題である。¹⁾

本報告では、医療情報部の仕事の中で大きな比重と意義を占めている、患者情報の2次活用に関する業務(研究支援、業務支援)の実情を報告し、特にDWHシステムにおける機能について実績とともにその機能分析と評価を行う。また、今後の医療情報分野の大きな課題の一つとして、非構造化データ(自然言語データ)からの知見や各種アウトカムの抽出があるが、退院時サマリーの文書データに対する取り組み事例についても取り上げる。

2. 患者情報の2次活用(DWH)システムについて

DWHシステムに関しては、前回のHIS更新時(2008年)にも構築、運用を試みたが、特にデータソースに対する定義・解釈等の問題により、当初の成果が挙げられなかった。このような経験を踏まえて、今回2015年のHIS更新においては、2段階でのデータ取り込みを計画した。まず、ある一定期間の実運用に必要なデータのみを先行取り込みを行い検証し、さらに、予定した全てのデータ収集、取り込みを行った。また、HISデータである電子カルテや医事会計からDWHへの直接インターフェースプログラムや抽出ツール(Ensemble)なども活用し、中間DBを用いることで効率を図った。その他の特徴は以下の通りである。

- HISベンダーが構築する電子カルテ、医事会計、看護支援等のみならず、部門システムとも連携を図り、DWHへの取り込み範囲を広げた。

- 文書管理システムとして新規導入するYahgeeとの連携を図り、テキストデータのみならずPDFファイルに存在するデータの2次活用を目指した。

構築したこのシステムは、処方情報、病名情報、検査結果などを活用し、継続的な条件を設定した構造化データへの対応を考慮した検索には適しているが、いわゆる自由文データである非構造化データへの対応については今後の課題である。そこで4章で、NLP(自然言語解析)システムの一つである、「iKnow」(インターシステムズ社提供)について触れる。

3. DWHシステムの実績と今後の課題

患者情報の2次活用としては、2つのアプローチで対応している。一つは、病院管理統計的な帳票の依頼であり、医事会計システムを用いてデータ抽出および帳票作成を行っている。医事会計システムには検査データが含まれていないため、検査データを含めた患者情報の抽出・統計業務依頼は、DWHから支援業務を行っている。医事会計システムおよびDWHからの患者データ抽出作業は、依頼票で管理されており、医療情報担当の事務職員3名が、他の業務とともに対応している。

3.1 実績とデータ抽出例

今年度の数か月分の実績を表1に示す。病院管理統計的な帳票の依頼が中心である、医事会計システムからのデータ抽出の方がDWHからの抽出に比べ、件数は多くなっている。今後、医学研究の活性化により、DWHからのデータ抽出件数の要望が増加することが予想されるが、対応職員のスキルとの関係もあり、迅速対応が今後の課題である。

表1 患者データ抽出件数

2007	4月	5月	6月	7月
DWH件数	22	17	19	22
医事会計件数	40	44	38	24
計	62	61	57	48



図1. 糖尿病患者が他疾患(糖尿病以外の疾患)で入院した場合、入院食事が常食となっており、糖尿病食となっていない事例が見受けられる為、適正な入院食事に改めることが目的とした患者抽出事例

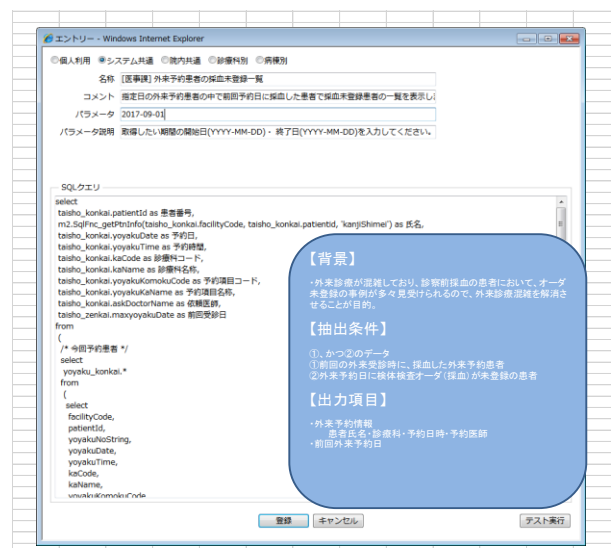


図2. 外来診療が混雑しており、診察前採血の患者において、オーダ未登録の事例が多々見受けられるので、外来診療混雑を解消させることが目的とした患者抽出事例

患者データ抽出事例を2例示したが、不適切な状況の分析、運用の改善に向けた取り組みが可能であり、医学研究支援のみならず、病院の運用改善への期待としても、実績を挙げていく必要がある。

3.2 DWHシステムの現状での問題点と要望

HISの更新は、2015年1月であった。DWHシステムにおいては、電子カルテシステム等からのソースデータの転送や

構築時のDBのバリデーションなどの準備作業のため、2015年5月頃から実質稼働してきている。約2年間の実績の中で、特に問題となっている点について、以下列挙する。

(1) DWHシステムにマスタ整備上の不備の問題

病棟、診療科、利用者などソースシステム上に存在するマスタの連携活用機能

(2) 患者指定のパラメータの柔軟性の問題

患者指定のデータを出力する場合、複数件指定や、一度の設定でして可能な機能(実態としてSQL記述にかなり時間を要し、プログラムが複雑化)

(3) 一時的テーブル(リスト等)作成機能の整備

本機能により、SQL記述が簡略化でき、労力削減が期待

(4) 外部ファイルとの連携機能の整備

外部ファイル(Excel 等)より取り込み機能の整備による、より柔軟な対処機能への期待

(5) 詳細検索画面(GUI)に対する機能強化

テーブル間の結合や件数集計などの簡易的な実行可能機能への期待

これらの問題は、HISベンダ、DWHベンダ、ユーザである医療情報部とのすり合わせ不足という側面もあるが、ユーザ側も十分に使いこなしていくことで、改善を図っていきたい。

4. 非構造化データ対策への試み

今後の医療情報分野の大きな研究課題の一つとして、非構造化データ(自然言語データ)からの知見や各種アウトカムの抽出がある。自然言語解析ツールの一つであるが、他分野ではすでに実績のあるインターシステムズ社の自然言語解析ツールである「iKnow」を取り上げ、プロトタイプ解析を紹介する。²⁾³⁾

4.1 特徴

通常のNLP(自然言語解析)システムは、事前に定義されたシソーラス、オントロジー、統計モデル等を活用したトップダウン方式であるに対し、iKnow はデータの長さや意味的複雑性に依存せず、ボトムアップ方式で重要な単語等のデータの指標化(インデックス化)を自動で行う、事前にシソーラスなどの情報を必要としないシステムである。

従来のNLPシステムで日本語を処理する際、形態素解析により文を単語に分割する手法が用いられるのが一般的である。しかし、この手法には(1)医療で用いられる専門用語や新出語を辞書に登録し継続管理する必要がある、(2)単語間の係り受けや結びつきは別のツールを用いて解析しなければならない、などの問題があった。これに対し iKnow は、文から意味的なつながりが強い単語を接続した「**エンティティ**」を検出する。例えば、形容詞＋名詞の接続が、単独の名詞より限定された意味を表すことから分かるように、エンティティは単独の単語に比べ、意味的な情報をより多く保持しており、文章の意味解析を行うための構成要素として適してとされている。

4.2 概要

日本語は英語などに比べ語順が柔軟であり、位置的に離れた単語同士が修飾関係にあることが多く、コンピュータによる意味解析を困難にしている。この問題に対し iKnow では、「**パス**」と呼ばれるエンティティの列を構成する。パスは、格助詞などを手掛かりとし、文に出現するエンティティを意味的な関連が強い順序に並べ替えたものである。

iKnow は、エンティティやパスを構成要素とし、文章の意味を定量的に評価するため、次のような指標を算出する機能を持つ。

○Frequency: エンティティが文書群中に出現する回数。

○Spread: エンティティが最低1回出現する文書数。

○ドミナンス(Dominance): ある文書におけるエンティティの「支配度」、すなわち、文書中の意味的な重要度。(エンティティを構成する単語の出現位置や出現頻度をもとに算出する。)

○近接度(Proximity): 2つのエンティティ間の関連性。(各パス内でのエンティティの出現位置により算出する。具体的には、あるパスにおける2つのエンティティの出現位置(先頭から1,2,...)の差の逆数を、すべてのパスについて和を取ったものである。)

4.3 退院時サマリーデータを対象にした評価

本院の電子カルテ上のフリーテキストデータのひとつである退院時サマリーデータを対象として解析を行った。退院時サマリーと電子カルテの特に経過記録(カルテ記事)に対して、iKnow を用いてエンティティを抽出し、両者のエンティティを用いたベクトル情報からコサイン類似度を算出した。具体的には iKnow の特性を利用して、以下の2点について解析を行った。

○退院時サマリーと対応するカルテ記事との類似度比較

○エンティティ間の類似度についての検討

今回対象としたのは、2年分 2307 件の退院時サマリーである。1番目の分析では、診療科ごとのデータを対象とした分析場合、全体として退院時サマリートのエンティティを軸とした比較の方が高い類似度を示している。これは、退院時サマリーがカルテ記事に記載されている内容のまとめであることを考えると、自然な結果と考えられる。2番目の分析では、今後のさらなる分析を待たなければならないが、1)言い換えや表記の揺れなどの曖昧性を考慮することによる意味解析・情報抽出の精度の向上、2)関連語などの発見的探索、3)シソーラス辞書や用語辞書の蓄積、などに応用可能ではないかと期待される。

5. まとめ

本報告では、長崎大学病院のDWHシステムによる患者情報の2次活用について紹介した。今後、システムの機能改善を図り、院内における2次活用が効率的に活用され、より有効な結果を算出していきたい。また、医療情報学研究では本稿で少し触れた、自然言語解析による文章データからの知識の抽出が課題である。また、プロセスマイニングによるログデータの解析環境の整備も大きな課題である。

謝辞

DWH構築、運用サポートについてはデータキューブ社小畑様、橋本様、また iKnow 関連についてはインターシステムズ社 Benjamin De Boe, 大平様、堀田様の多大なご尽力に感謝申し上げます。

参考文献

- 1) Masayuki HONDA, Takehiro MATSUMOTO, A new Hospital Information System Toward a promotion for security and BCP strategy, Digital Healthcare Empowering Europeans R. Cornet et al. (Eds.); MIE-2015, 962-963, 2015
- 2) 本多 正幸, 松本 武浩, 浅田 真瑞, 牛嶋 拓也, 小畑 恭弘, 橋本 哲朗, 堀田 稔, 植松 裕史, 自然言語解析ツール「iKnow」を用いた退院時サマリーデータの評価ー長崎大学病院DWHを対象とした取組ー, 医療情報学, 36(Suppl.), 1050-1053, 2016
- 3) Masayuki Honda, Takehiro Matsumoto A new essential functions installed DWH in Hospital Information System -Process Mining techniques and Natural Language Processing are focused on-, MEDINFO2017, (24-Aug-17 Board#24 Submission#47), 2017