

一般口演 | 医療データ解析

一般口演25

医療データ解析

2019年11月24日(日) 13:40 ～ 15:10 F会場 (国際会議場 3階中会議室302)

[4-F-3-04] 多施設テキストデータベースを用いた退院時サマリー作成・監査支援の試み

○鈴木 隆弘¹、土井 俊祐²、木村 隆¹、島井 健一郎¹、嶋田 元³、畠山 豊⁴、松村 泰志⁶、松本 武浩⁵、横井 英人⁷、本多 正幸⁵（1. 千葉大学医学部附属病院企画情報部, 2. 東京大学医学部附属病院企画情報運営部, 3. 聖路加国際病院医療情報センター, 4. 高知大学医学部附属医学情報センター, 5. 長崎大学医学部附属病院医療情報部, 6. 大阪大学大学院医学系研究科医療情報学, 7. 香川大学医学部附属病院医療情報部）

キーワード：Teximining, Discharge summary, Multi institutional joint research, BERT

我々はこれまでに千葉大学医学部附属病院、聖路加国際病院、高知大学医学部附属病院、長崎大学病院、大阪大学医学部附属病院、香川大学医学部附属病院の6施設から約32万例の退院サマリーを収集して解析を行い、大規模なテキストデータベースを構築すると共に、これを用いた応用アプリケーションの開発に取り組んできた。本年は昨年開発した、退院期間を指定して複数のサマリーを一括でチェックするアプリケーションを改良して評価を行ったので報告する。

本アプリケーションは電子カルテから退院サマリーのテキスト部分を読み込み、構築済みのテキストデータベースとの照合をベクトル空間モデルを応用して行い、推定されたDPCコードを付与して出力する。退院期間、診療科、主治医、参照するデータベースの施設等を指定して分析を行うことが可能である。汎用性を高めるためにWEBアプリケーションとして開発しており、クライアント側は一般的なWEBブラウザが利用出来る。演題申込みの投稿時点では分析の途中であるが、診療情報管理士による診療録質的監査の支援、医事課によるDPCコードチェック、研修医による退院時サマリー作成時の支援などに応用して、評価を行っている。本年度から新たに群馬大学医学部附属病院の参加を得て、テキスト解析に新しい手法を導入し、従来の方法との比較も実施する予定である。

多施設テキストデータベースを用いた退院時サマリー作成・監査支援の試み

鈴木隆弘^{*1}、土井俊祐^{*2}、近本圭祐^{*3}、木村隆^{*1}、島井健一郎^{*1}、嶋田元^{*4}、野口令^{*5}、畠山豊^{*6}、
本多正幸^{*7}、松村泰志^{*8}、松本武浩^{*7}、横井英人^{*9}、川上英良^{*3}

*1 千葉大学医学部附属病院企画情報部、*2 東京大学医学部附属病院企画情報運営部、
*3 千葉大学医学部人工知能(AI)医学、*4 聖路加国際病院医療情報センター、
*5 群馬大学医学部附属病院システム統合センター、*6 高知大学医学部附属医学情報センター、
*7 長崎大学病院 医療情報部、*8 大阪大学大学院医学系研究科医療情報学、
*9 香川大学医学部附属病院医療情報部、

Development of discharge summary audit support application by text mining

Suzuki Takahiro^{*1}, Doi Shunsuke^{*2}, Chikamoto Keisuke^{*3}, Kimura Takashi^{*1}, Shimai Ken-ichiro^{*1}, Shimada Gen^{*4},
Noguchi Rei^{*5}, Hatakeyama Yutaka^{*6}, Honda Masayuki^{*7}, Matsumura Yasushi^{*8}, Matsumoto Takehiro^{*7},
Yokoi Hideto^{*9}, Kawakami Eiryo^{*3}

*1 Department of Medical Informatics and Management, Chiba University Hospital
*2 Department of Healthcare Information Management, The University of Tokyo Hospital
*3 Department of Artificial intelligence medicine, Chiba University
*4 St. Luke's International Hospital Medical Information System Center
*5 Gunma University Hospital System Integration Center
*6 Center of Medical Information Science, Kochi University
*7 Department of medical informatics, Nagasaki University Hospital
*8 Medical Informatics, Osaka University Graduate School of Medicine
*9 Department of Medical Informatics, Kagawa University Hospital

Abstract in English comes here.

We have collected and analyzed about 320,000 discharge summaries from 6 institutions, Chiba University Hospital, St. Luke's International Hospital, Kochi University Hospital, Nagasaki University Hospital, Osaka University Hospital and Kagawa University Hospital. We have built a large text database from these and have been working on application development. We will report on the improvement and evaluation of the application developed last year. This application extracts the text part of the discharge summary from the electronic medical record and collates it with a text database by applying a vector space model. The keywords in the sentence, the estimated DPC code, ICD10 code, and 10 keywords used as the basis for selection are output. Analysis can be performed by specifying the discharge period, patient ID number, and the facility of the database to be referenced. It was evaluated by users that it was useful for pre-checking long summary audits, leading to a reduction in coding time, but further improvement in response was required. Since this year, Gunma University Hospital has been newly joined, so we will compare and cross-match with other hospitals. Furthermore, we plan to introduce the BERT method, which is a kind of deep learning, in text analysis, and report the results of comparison with conventional methods.

Keywords: Please input 3 to 5 keywords in English.

1. はじめに

我々はこれまでに千葉大学医学部附属病院、聖路加国際病院、高知大学医学部附属病院、長崎大学病院、大阪大学医学部附属病院、香川大学医学部附属病院の6施設から約32万例の退院時サマリーを収集して解析を行い、大規模なテキストデータベースを構築すると共に、これを用いた応用アプリケーションの開発に取り組んできた。本年は昨年報告した退院サマリーを分析するアプリケーション¹⁾を改良して評価を行ったので報告する。また、群馬大学医学部附属病院が新たに参加すると共に、単語の重み付けに BERT (Bidirectional Encoder Representations from Transformers)を取り入れ、従来の方法との比較を開始した。

2. 対象

千葉大学病院、長崎大学病院、香川大学病院、大阪大学

病院、高知大学病院、佐賀大学病院、聖路加国際病院の電子カルテに対応した退院時サマリー収集と解析のためのプログラムを作成し、2012-2015 年の 4 年間に入院した症例を対象として、サマリー本文のテキストと 14 桁の DPC 情報の収集を行い、解析のためのモデルデータを作製した。データの収集と基本的な処理は各施設内で行い、形態素分解後の連結不可能匿名化した情報のみを千葉大学に集約した。

対象となった退院時サマリー数は千葉大学 67266 例、大阪大学 58726 例、香川大学 20836 例、高知大学 34994 例、長崎大学 61582 例、聖路加国際病院 73065 例で、総計は 316459 例に達した。群馬大学医学部附属病院については既に約 3 万例の抽出が終了しており、本稿執筆時は倫理審査等の手続き中であるが、大会時には分析結果を発表できる見込みである。

3. 方法

3.1. 形態素解析とベクトル化

抽出したサマリーから重要語を抽出してテキストデータベースを作成した。従来は形態素解析に **MeCab** を使用し、解析用辞書としては、我々がこれまでに作成した23万語の医療用語辞書とパラメディカル用医療辞書の **ComeJisyo** を併用した。得られた索引語の重み付けには **tf×idf** 法を用いてベクトル空間を構築した。本年は元のテキストを保存していた千葉大のサマリーに対して **BERT** による解析を行い、従来の方法との比較検討を施行中である。

3.2. DPC 自動判定と病院間クロスマッチ

昨年の発表と同様に各施設から症例数が 20 以上の DPC
 コードを抽出し、7 対 3 の比率で無作為にモデル作成用と検
 証用のサマリーに分け、検索対象のサマリーベクトル空間と
 入力されたサマリーを比較して最も類似したベクトルを検索し
 DPC を判定する。参加 7 病院各々の検証用データとモデル
 データおよび、全病院のデータを統合したモデルデータとの
 間で 7 対 8 のクロスマッチ判定自動判定を、DPC コードの 14
 桁全てと上位 6 桁のそれぞれで行う。

3.3. Web アプリケーション

昨年発表した Web アプリケーションを改良し、患者を直接指定して類似度の高いサマリーをキーワードと共に提示する機能を追加した。アプリケーションの開発にはM言語データベースの Cache を用いている。(図1)。

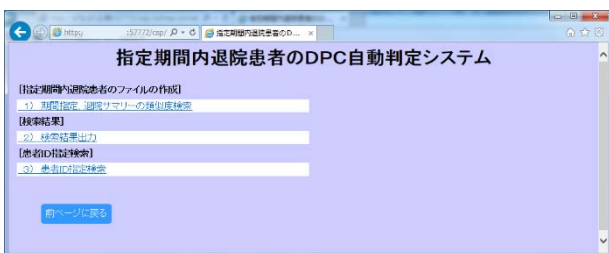


図1. システムメニュー画面

患者 ID 番号を入力し、照合対象とする病院別モデルデータを指定すると入院歴の選択画面が表示され、対象のサマリーを選択することで解析を開始する。



図2. 類似サマリー検索システム結果表示画面

解析はサマリー単位で行われ、類似度の高い順に 10 例の DPC コード、ICD10 コード、原文へのリンク及び重み大きな順に 10 個のキーワードを表示する。また、画面の上部にはサマリーを形態素解析した索引語の一覧を表示している。(図2) アプリケーションの精度と使い勝手の検証のために、診療情報管理士と総合医療教育研修センターによる評価を行った。

4. 結果

診療情報管理士からは、従来は期間指定だったために多数の患者を処理することから、1日分であっても30分以上の時間を要して使いづかったが、直接患者を指定することで使いやすくなったと評価された。しかしながら1例の処理に30秒以上かかるため、レスポンスの一層の向上が求められた。索引語が列挙されているので、長文サマリーの診療科の監査に当たっての事前チェックには有用で、コーディング作業の時間短縮に繋がると評価された。医師からも索引語の一覧は経過をまとめる意味で有用であり、いろいろな候補が表示されるので、コーディングの参考になる、研修医本人よりもアテンディングドクターが指導をする際の参考に役立つ、等の評価が得られた。

他院のモデルデータを適用した場合、判定された DPC コードが同じ症例でも、キーワードはかなり異なっていることがあり、施設による用語の差が大きいことを示唆した。(図3)

[illegible]

図3. 他院モデルデータによる判定結果

5. 考察

昨年の評価を踏まえて Web アプリケーションの改良を行い、一定の評価は得られたが、レスポンス・精度ともに充分とは言えず、常時実用とするには至らずに、参考資料としての利用に留まっている。そのため、精度を高める目的で BERT の利用を試みた。BERT は 2018 年に Google が発表した自然言語処理モデルで、特徴として汎用性の高さが挙げられるだけでなく、今後「文脈の理解」や「暗黙の了解」など、より深いレベルでの言語理解に役立つと考えられている。しかし日本語の医療テキストへの応用例は未だ少なく、今回の試みは先駆けとなるものであり、大会時には従来法との比較結果を提示できる見込みである。

参考文献

- 1) 鈴木隆弘, 土井俊祐ら. 退院サマリー監査を支援するDPC判定アプリケーション. 第38回医療情報学連合大会論文集 2018; 786-787)