

平成 30 年 6 月 11 日現在

機関番号：16401

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K08847

研究課題名(和文) 病院情報システムにおける病名の信頼性評価方法の確立

研究課題名(英文) Establishment of reliability evaluation method of disease name in hospital information system

研究代表者

奥原 義保 (OKUHARA, YOSHIYASU)

高知大学・教育研究部医療学系連携医学部門・教授

研究者番号：40233473

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：病院情報システムのデータから病名の信頼度を求めることを目指した。病院情報システムの確定病名につき、症例群と対照群で検査、処方、処置を実施している割合の比を求め、値が大きい項目を選択、樹木モデルによりルールを求めた。また、モデルに取り入れた項目がどのような病名に対して実施されているかの割合を求め、その項目を特徴づける特徴量としてクラスター法で分類することにより、真の病名を識別することができた。また、求めたルールを検証データに適用し、オーダデータのみである程度の精度が得られることがわかった。電子カルテの記述部分を情報として使うためには、潜在トピックモデルを適用するのが有効であることがわかった。

研究成果の概要(英文)：We aimed to obtain reliability of disease name from data of hospital information system. For the definite disease name of the hospital information system, the proportion of examination, prescription and treatment being conducted in the case group and the control group was obtained, the item with the large value was selected, and the rule was obtained by the decision tree model. In addition, it was possible to identify the true disease name by finding the ratio of what kind of item taken in the model to the name of disease and classifying it by the cluster method as the characteristic quantity characterizing the item. Moreover, we applied the obtained rule to the verification data and found that some degree of accuracy can be obtained only with the order entry data. To use the narrative part of the electronic medical record as information, it turned out that applying a latent topic model is effective.

研究分野：医療情報学

キーワード：病院情報システム 機械学習 病名評価

1. 研究開始当初の背景

我が国において、初めて総合的な病院情報システムがトータルオーダエントリシステムとして導入されてから既に 33 年、初めて電子カルテが導入されてから 14 年が経過している。この間、多くの病院にオーダエントリシステムや電子カルテが普及し、日々の診療に伴う膨大なデータが蓄積されている。病院情報システム構築の当初の目的の中には、システムに蓄積された医療データを医学の発展のために有効に活用することが含まれており、大きな期待を担っていたはずである。

しかしながら、そうしたデータは、各施設の運営や経営のためにある程度活用されているものの、医学の発展のために十分活用されるという使い方とは程遠い使われ方しか実現していない。その最も大きな原因の一つとして、病院情報システムに登録される病名の信頼性が明確でないことが挙げられる。

申請者らが所属する高知大学医学部では、1981 年に我が国において初めてとなるトータルオーダエントリシステムを自主開発・導入し、運用を開始した。以来、患者年齢、性別、来院日、入院日、病名、検査項目、検査結果、処方歴、注射実施歴、処置実施歴、手術歴等のデータをすべてデータウェアハウス(DWH)化し、保存してきた。2010 年には、これらのデータをハッシュ関数を用いた不可逆的変換によって非連結匿名化し、日付情報も患者毎に生年月日からの相対日付に変換、さらに DWH の利用環境を仮想システム化するという、徹底した個人情報保護とデータセキュリティ対策を施したシステム RYOMA(Retrieval sYstem for Open Medical Analysis)も構築した。申請者らは、これらのシステムのデータを医学研究や教育へ活用する試みを継続しており、成果を出しつつある。

しかしながら、病院情報システムのデータが、保険医療機関の日々の診療行為に基づいているという性格上、一部のデータの信頼性には課題がある。その最大の課題は病名の信頼性である。病院情報システムに登録された病名には、診断確定前の疑い病名が転帰されずに残っている場合や、効能が知られていながら保険適用の実現が間に合っていないために、止むを得ずつける病名などが含まれていると考えられる。

こうした病名の信頼性の問題は、今後盛んになることが予想されるレセプトデータを統合したビッグデータの解析においても課題となるはずである。また、レセプトデータを統合したビッグデータの解析の次には、より情報量の多い個々の病院情報システムのデータを統合した National Database の段階へと進むと考えられるが、病名の信頼性の解決なくしてはビッグデータの解析も意味を持たないと考えられる。このように、病院情報システムにおける病名の信頼性は、病院情報システムのデータやレセプトデータが真

のビッグデータ解析に使えるかどうかも左右する重要な要素であり、喫緊に解決が求められる課題である。こうした病名の信頼性の問題を解決する試みは、レセプトデータを対象としていくつか始まっているが、本研究では、より情報量の多い病院情報システムのデータによって、より正確に信頼度を定めることを目指している。

2. 研究の目的

申請者等は単一施設内ではあるが、1981 年以來のオーダエントリシステムのデータおよび 2007 年以來の電子カルテの記載情報を用いて、病名の信頼度を定量的に決める方法を確立することを目指す。そのため、本研究では、以下の 2 点を目的とする。

(1) 病院情報システムで登録された入院患者の確定病名につき、検体検査、処方、処置、退院サマリの情報を用いて病名に至るルールを導き出し、そのルールに含まれることが考えられる真の病名と、その他の疾患の治療を目的として付けられた病名を識別する方法を考える。

(2) また、そうしたルールを正解のわかっている検証データに適用し、どのくらい正確に正解が判別できるかを確認する。

以下(1)(2)という項目番号を、これらの目的に対応した記述の識別のために使用する。

3. 研究の方法

【対象患者】

高知大学医学部附属病院の総合医療情報システム I M I S に蓄積されたデータを匿名化した RYOMA 上の、

(1) 2002 年 3 月以降の全患者

(2) 2003 年 6 月以降の入院患者。

を対象とする。

【オーダ情報を用いた病名判別のルール導出手順】

(1)(2) 共通

対象となる確定病名が登録されている群を症例群、対象となる確定病名はもちろん、疑い病名も登録されていない群を対照群とする。

症例群では、対象病名登録日の前後 30 日以内の検査値について平均値を計算、さらに対象病名登録日の前後 30 日以内の処方の有無、対象病名登録日の前後 30 日以内の処置の有無を求めた。対照群では、各患者の各検査の最終日以前 30 日以内の検査値の平均値、全期間での処方の有無、全期間での処置の有無を求めた。続いて、症例群と対照群の各群で検査、処方、処置を実施している割合を求め、さらに両群での割合の比(症例群での割合/対照群での割合)を求めた。

各分類(検査、処方、処置)において統計的に有意な比のうち値が大きい方から(症

例群に特異的な) 10 項目のみを選択。

で作成したデータに対して、対象確定病名の有無を目的変数、各分類(検査、処方..等)を説明変数として CART に基づく樹モデルによる解析を行った。解析に用いたソフトウェアは R 3.0.0 パッケージ rpart である。

さらに、(1)に関しては、全体での頻度は低くても、高い頻度で当該確定病名に結びついているパターンを検出するため、作成したデータに対して、選択項目をアイテムとし、確定病名の有無を結論部とした association 分析も行う。その際、Support (指示度)は低く設定する(0.01 以上程度)、Confidence は大きく設定する(0.5 以上程度)。解析に用いたソフトウェアは R 3.0.0 パッケージ arule である。特定したパターンをルールと考える。

(1)について得られるルールには、真の病名と、他の疾患に対する治療上の必要性から登録された病名の両方が含まれているはずである。これらのパターンを判別するため、処方の分類で対象病名に特異的であるとしてリストアップされた薬剤 20 項目について分類を試みた。薬剤の分類にあたって、薬剤マスターの薬効分類や成分・機序の情報を用いるのではなく、どのような病名に対して使われているかを実際にデータから求めることによって行った。具体的には、次のような手順に従った。

まず、特定の薬剤につき

・当該薬剤の最初の処方日前後 30 日以内に登録された病名を列挙

- ・用いた病名分類の粒度は ICD の大分類 (ただし、検診や事故などに関係した病名分類である先頭が 'X', 'Y', 'Z', 'U', 'V' の病名は、は除いた。)
- ・この粒度での全病名分類からなるベクトル空間を考える。
- ・各病名分類の頻度を求める
- ・特定の薬剤に対する病名分類の全頻度に対する各病名分類の割合を求める。
- ・この割合をその薬剤を特徴づける特徴量(ベクトル)と考える。
- ・これら特徴量を階層型クラスター法によって分類する。

逆流性食道炎の病名が付いた患者とついていない患者に対する処方の有無の違い(両群での実施割合の比で評価)が大きかった薬剤上位 20 位までを対象に上の方法で特徴づけを行った。

【(2)の解析】

(1)で対象とした逆流性食道炎の病名は、DPC 主病名としては症例数が少なかったため、本研究での方法の妥当性を感度・特異度によって評価する解析は、本院での DPC 主病名として症例数が多い「肝細胞癌」を対象に行った。

まず肝細胞癌という確定病名が登録されているが、DPC 病名としては肝細胞癌が登録されていないグループを抽出した。このグループを症例群とし、肝細胞癌という病名が疑い病名としても登録されていないグループから半数をランダムに抜き出したグループを対照群として学習用データを作成し、(1)と同様に樹モデルを用いてルールを導き出す。

続いて、DPC 主病名として肝細胞癌が登録された症例を症例群(この群は真の病名がつけられた群と考えられる)とし、疑い病名としても登録されていないグループからランダムに半数を抜き出した残りの半数を対照群として、これらから成るグループを検証の対象である検証データとした。学習用データから導き出したルールによって、検証データの正解がどの程度正しく判別できるかは、R 3.0.0 パッケージ rpart の関数 predict によって計算する。

【オーダ情報に退院サマリの情報を加えた病名判別のルール導出手順】

退院サマリの記述から形態素解析により抽出した形態素を各サマリの出現順序数及び詳細品詞情報を保持させたままデータベース化した匿名化 DWH より、検査や処方の場合と同様に、対象病名が登録されている場合とそうでない場合の各形態素に対する存在割合の比を計算して、値の高いものを選択し、オーダデータだけからなるデータに追加し樹モデルの解析を行う。

4. 研究成果

【結果】

(1) **逆流性食道炎を対象とした真の病名と、その他の疾患の治療を目的として付けられた病名を識別するための解析**では、確定病名有に特異的な項目は、**検査関連**：総ビリルビン、ALT、AST、クレアチニン、赤血球数、ヘマトクリット、ヘモグロビン、平均赤血球量、血小板数、白血球数、**処方関連**：フォスマック錠 35mg、ランソプラゾール OD 錠 15mg、ネキシウムカプセル 20mg、ネキシウムカプセル 10mg、オメプラール錠 20、パリエット錠 10mg、タケプロン OD 錠 15、タケプロン OD 錠 30、プリドール注射用 500、タケプロン静注用 30mg、**処置関連**：心電図検査、血圧・脈波検査、単純一般、血管造影、一般 CT、レスピラトル使用、鼻腔栄養・経管栄養、ECG 12、日赤血赤血球液 LR、日赤血新鮮凍結血漿 LR のようになった。

これらを説明変数に用いた樹モデルによる解析の結果、次のようなルールが導き出された。

1. タケプロン OD 錠 15(プロトンポンプ阻害薬(PPI))あり AND 心電図検査なし => (逆流性食道炎の病名あり：0.85, 逆流性食道炎の病名なし 0.15)
2. タケプロン OD 錠 15 (15mg) (プロトン

ポンプ阻害薬 (PPI)) あり AND 心電図検査あり AND 血管造影 なし => (逆流性食道炎の病名あり : 0.68, 逆流性食道炎の病名なし 0.32)

- 3 . タケプロン OD錠 15(プロトンポンプ阻害薬 (PPI)) なし AND パリエット錠 10mg (プロトンポンプ阻害薬 (PPI)) あり AND 心電図検査なし => (逆流性食道炎の病名あり : 0.79, 逆流性食道炎の病名なし 0.21)
- 4 . タケプロン OD錠 15(プロトンポンプ阻害薬 (PPI)) なし AND パリエット錠 10mg なし AND タケプロン OD錠 30 なし AND ネキシウムカプセル 20mg (プロトンポンプ阻害薬 (PPI)) あり AND 心電図検査なし => (逆流性食道炎の病名あり : 0.79, 逆流性食道炎の病名なし 0.21)
- 5 . タケプロン OD錠 15(プロトンポンプ阻害薬 (PPI)) なし AND パリエット錠 10mg なし AND タケプロン OD錠 30 なし AND ネキシウムカプセル 20mg (プロトンポンプ阻害薬 (PPI)) なし => (逆流性食道炎の病名あり : 0.03, 逆流性食道炎の病名なし 0.97)

さらに、同じデータを用いた association 分析の結果、Support (指示度) は低くても Confidence (確信度) の大きいパターンとして次のようなルールが見出された。

タケプロン OD錠 15 とタケプロン OD錠 30 とフォサマック錠 35mg (骨粗鬆症治療薬) が同時に処方されている場合 (指示度 : support=0.1%) => 93%の確率 (確信度 : confidence) で逆流性食道炎の病名あり
この結果は、全体での頻度は低いものの、このパターンに対しては高い頻度で逆流性食道炎という病名がつけられていることを表す。

これらのルールには、真のルールと、他の疾患に対する治療上の必要性から登録された病名のルールの両方が含まれているはずである。これらのパターンを判別するため、処方の分類で対象病名に特異的であるとしてリストアップされた薬剤 20 項目について階層型クラスター分類を実施した結果、薬剤のクラスターは、大きく抗凝固剤・抗血小板剤、骨粗鬆症治療薬・合成副腎皮質ホルモン剤、(プロトンポンプ阻害薬 (PPI))・プロトンポンプ阻害剤 (H2 ブロッカー)・消化管運動機能改善薬、その他点滴に関連した薬剤の 4 つに分かれた。

(2) 肝細胞癌の学習用データを対象とした、導き出されたルールを正解のわかっている検証データに適用し、どのくらい正確に正解が判別できるかを確認した解析では、確定病名有に特異的な項目は、**検査関連** : PIVKA-(EIA)、PIVKA(RIA)、PIVKA2、AFP レクチン反応性分画、AFP レクチン反応性 L1 分画、AFP レクチン反応性 L2 分画、AFP レクチン反応性

L3 分画、ICG 負荷前採血、ICG 負荷 15 分後残存率、**処方関連** :

エタノール、ゼルフォーム No.12、ラクツロース・シロップ 60%「コーワ」、リーバクト顆粒、アミノレバン点滴静注、ファルモルピシン注、リピオドール 480 注、オプトレイ 350 注、オピスタン注射液、ソナゾイド注射用、**処置関連** : 血管造影、エタノール局所注入、(クールチップ) ラジオ波焼灼法、検温、血圧測定、蓄尿、尿比重、安静度、清潔、洗髪

これらを説明変数に行った樹モデルによる解析の結果、導き出されるルールのうち主なものは以下のようであった。

- 1 . リピオドール 480 注 (血管内塞栓促進用補綴材) あり AND ファルモルピシン注 (抗悪性腫瘍薬) あり => (肝細胞癌の病名あり : 0.97, 肝細胞癌の病名なし 0.03)
- 2 . リピオドール 480 注 (血管内塞栓促進用補綴材) あり AND 血管造影 あり => (肝細胞癌の病名あり : 0.77, 肝細胞癌の病名なし 0.23)
- 3 . リピオドール 480 注 (血管内塞栓促進用補綴材) あり AND 血管造影 なし AND ICG 負荷 15 分後残存率 15%以上 => (肝細胞癌の病名あり : 0.94, 肝細胞癌の病名なし 0.06)
- 4 . リピオドール 480 注 (血管内塞栓促進用補綴材) なし AND 血管造影 なし AND ICG 負荷 15 分後残存率 22%未満 => (肝細胞癌の病名あり : 0.008, 肝細胞癌の病名なし 0.92)

導き出されたルールを検証データに適用して求めた感度は 0.67 特異度は 0.88 であった。

【退院サマリの情報を用いた解析】

実際に肝細胞癌を対象に各形態素に対する存在割合の比を計算したところ、上位に登場したのは、lipiodol、HCC、epirubicin、PEIT、肝細胞癌、CTAP、RFA、TAE、C 型慢性肝炎、epirubicine のような単語であった。

これらはいずれも既にオーダデータだけからなるデータに表れている情報または病名そのものであり、新たなルールを付け加える可能性はない。このため、これらのデータを加えてオーダデータだけからなるデータに追加し樹モデルの解析を行うことは行わなかった。

そのかわり、潜在トピックモデルの適用を検討し、効果を検証するために、初診時記録及びオーダ情報によって分割された患者集団における各病名の登録率の評価を行った。初診時記録に対して潜在トピックモデルによって計算された各トピックに対する確率及びオーダの有り無しによる距離に基づき、それぞれ階層化クラスタリングを実施し、細分化されたクラスタ毎の登録率を取得した。

表 1: 初診時記録に基づくクラスタリング

初診時クラスタ	オーダクラスタ	患者数	ICD10	ICD10登録患者数
1	1	13	A09	10
1	6	4	A09	3
2	2	40	I10	37
3	3	18	F48	14
3	2	17	F41	14
4	3	57	R68	11
5	2	18	M54	6
5	3	9	M54	6
6	3	15	J00	6
6	8	14	J06	6
7	4	9	F32	6
7	2	5	F32	2
8	2	30	I10	27
9	2	13	R42	10
9	3	8	R51	4
10	2	10	R63	10
11	2	19	M47	17
12	2	9	T78	10
12	7	7	A37	7
13	2	20	R10	17
14	2	23	M62	20

表 2: 初診時記録及びオーダ情報に基づくクラスタリング

初診時クラスタ	人数	トピック	Rank1	Rank2	Rank3	Rank4	Rank5
1	20	12	下痢	°C	腹痛	食べる	嘔吐
2	46	11	いる	こと	できる	本人	られる
3	64	5	仕事	こと	する	ストレス	不安
4	75	13	舌	便秘	腹部	脈	食欲
5	36	10	痛み	痛い	右	前	整形
6	64	3	咳	°C	痰	発熱	処方
7	21	18	人	いる	思う	歳	子供
8	56	6	する	内科	上	程度	認める
9	31	19	頭痛	めまい	感じ	肩こり	ない
10	18	15	いる	とき	する	学校	食べる
11	37	14	受診	こと	紹介	相談	希望
12	18	4	こと	処方	喫煙	飲酒	特記
13	28	20	受ける	希望	来院	not	診察
14	29	2	圧痛	可能性	貧血	出現	受診

高知大学医学部附属病院総合診療部受診の543人のデータに適用した結果、処方歴がある患者数5人以上のクラスタでの最大病名登録率は77%となった。

【考察】

(1) 逆流性食道炎を対象とした真の病名と、その他の疾患の治療を目的として付けられた病名を識別するための解析では、

逆流性食道炎の病名に特異的な項目では、検査項目には一般的な検査項目が現れ、特に特徴のあるパターンとは言えない。一方で処方項目にはプロトンポンプ阻害剤やH2ブロッカーなど、逆流性食道炎の治療に用いられる薬剤が多く表れた。同時に、アレンドロン酸ナトリウムであるフォサマック錠 35mg や 合成副腎皮質ホルモン剤であるプリドール注射用 500 (500mg) なども現れた。

これらの薬剤に対して、どのような病名に対して使われているかを、これらの薬剤が実際にどのような病名にどのくらい使われているかの頻度を特徴ベクトルとして階層型クラスタ分析で分類した結果、大きく1. 抗凝固剤・抗血小板剤、2. 骨粗鬆症治療薬・合成副腎皮質ホルモン剤、3. (プロトンポンプ阻害薬 (PPI))・プロトンポンプ阻害剤 (H2 ブロッカー)・消化管運動機能改善薬、4. その他点滴に関連した薬剤の4つに分かれた。

本来、逆流性食道炎に対しては、3番目の

クラスターに属する薬剤だけが特異的に使用されるはずであるが、1番目、2番目のクラスターに属する薬剤も特異的に使用されている。1番目、2番目のクラスターに属する薬剤に消化管潰瘍を生ずる恐れが言われており、プロトンポンプインヒビターやH2ブロッカーはそれに対する予防あるいは治療のために使用された可能性がある。

実際に、選択項目をアイテムとし、確定病名の有無を結論部とした association 分析を行うと、PPI や H2 ブロッカーとともに骨粗鬆症の治療薬を処方するパターンが support (指示度) は低いものの Confidence が大きい結果に表れてくる。つまり、全体における頻度は低いものの、これらのパターンに対しては高い頻度で逆流性食道炎という病名がつけられていることを表す。

なお、処置の項目については、循環器に関連した生理検査の項目が多く見られた。これは、逆流性食道炎の症状には循環器系の疾患と共通したものが多いため、鑑別のために行われたことが反映していると推測される。

樹モデルを用いたルールの抽出では、プロトンポンプ阻害薬 (PPI) の有無や心電図検査実施の有無などが逆流性食道炎の有無に影響の大きいルールであることがわかった。また、association 分析の結果からは、全体での頻度は低いものの、プロトンポンプ阻害剤と骨粗鬆症治療薬の同時処方というパターンに対しては高い頻度で逆流性食道炎という病名がつけられていることがわかった。

(2) 肝細胞癌の学習用データを対象とした、導き出されたルールを正解のわかっている検証データに適用し、どのくらい正確に正解が判別できるかを検証した解析では、検査、処方、処置とも、肝細胞癌に特異的な項目が現れた。

導き出されたルールを検証データに適用して求めた感度は0.67 特異度は0.88 という特に高い値ではなかったが、ルールの導出に際して、逆流性食道炎の際のように、真のルールと、他の疾患に対する治療上の必要性から登録された病名のルールを識別する解析を行っていないこと、電子カルテの記述の情報を使っていないことなどからこうした結果になったと考えられる。

退院サマリの情報を用いた解析では、検査や処方、処置などの項目で対象病名に特異的な項目を抽出したのと同じ方法では既にオーダデータだけからなるデータに表れている情報以上の情報は得られなかったが、潜在トピックモデルを用い、初診時記録及びオーダ情報によって分割された患者集団における各病名の登録率の評価を行った結果からは、特徴的な記述及び処方歴の集団では類似した背景を持つ患者群が抽出可能なことを確認し、特徴的なパターンのルールを導出するための手法として有望であることを示した。

5. 主な発表論文等
(研究代表者、研究分担者及び連携研究者には下線)

(3) 連携研究者 なし

(4) 研究協力者 なし

[雑誌論文](計 0件)

[学会発表](計 1件)

発表者名: 畠山 豊, 永田 桂太郎、渡部 輝明、奥原 義保

発表タイトル: 初診時記録及びオーダ情報に基づく病名登録率評価

学会等名: 第22回日本医療情報学会春季学術大会(シンポジウム2018 in 新潟)

発表年: 2018年

[図書](計 0件)

[産業財産権]

出願状況(計 0件)

取得状況(計 0件)

6. 研究組織

(1) 研究代表者

奥原 義保 (OKUHARA, Yoshiyasu)

高知大学・教育研究部医療学系連携医学部門・教授

研究者番号: 40233473

(2) 研究分担者

畠山 豊 (HATAKEYAMA, Yutaka)

高知大学・教育研究部医療学系連携医学部門・准教授

研究者番号: 00376956

渡部 輝明 (WATABE, Teruaki)

高知大学・教育研究部医療学系連携医学部門・講師

研究者番号: 90325415

中島 典昭 (NAKAJIMA, Noriaki)

国立研究開発法人国立がん研究センター・情報統括センター・研究員

研究者番号: 00335928

片岡 浩巳 (KATAOKA, Hiromi)

川崎医療福祉大学・医療技術学部・教授

研究者番号: 80398049

寺田 典生 (TERADA, Yoshio)

高知大学・教育研究部医療学系臨床医学部門・教授

研究者番号: 30251531

堀野 太郎 (HORINO, Taro)

高知大学・教育研究部医療学系臨床医学部門・講師

研究者番号: 90448382