

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 19 日現在

機関番号：17701

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26460868

研究課題名(和文) データマイニング技法を応用した医療の質とコストの適正化要因の発見手法に関する研究

研究課題名(英文) Study on finding method of factor of medical quality and cost appropriate for data applying data mining technique

研究代表者

熊本 一郎 (Kumamoto, Ichiro)

鹿児島大学・医歯学域医学系・教授

研究者番号：40225230

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)：本研究はDPC別医療コスト分析用DWHを構築し、データマイニング技法を用いて、医療コストに影響を与える因子発見に適した手法の評価を目的とした。2011年～2015年の肝細胞癌の初回入院・初回治療の365例について分析した。その結果、k-mean法によるクラスタリングでは、在院日数のみがコストに関連していた。高コストとなる費目分析については、ニューラルネットワークによる機械学習のみ分析可能であった。データを機械学習用の274例と評価用の91例に分割し、機械学習したモデルで、評価用の費目から、高コストになるか否かを判定させた。その結果、80%以上の確度で費目名から高コストと判断することができた。

研究成果の概要(英文)：The purpose of this study was to construct a DWH for medical cost analysis by DPC, and to evaluate a method suitable for finding a factor affecting medical cost by using data mining technique. We analyzed 365 cases of initial hospitalization and initial treatment of hepatocellular carcinoma from 2011 to 2015. As a result, in clustering by the k-mean method, it was found that only the number of hospital days is related to cost. With regard to the cost analysis, which is a high cost, only machine learning by a neural network could be analyzed. We divide the data into 274 cases for machine learning and 91 cases for evaluation, and in the machine learning model, we judged from expenditure items for evaluation whether or not the cost becomes high. As a result, it was possible to judge from cost item name to high cost with accuracy of 80% or more.

研究分野：医療情報学

キーワード：病院データウェアハウス 医療コスト分析 データマイニング ニューラルネットワーク

1. 研究開始当初の背景

少子高齢化が進行中の我が国では、厚生労働省によって様々な医療コスト抑制策が実施されている。しかしながら、総医療費の増大にはなかなか歯止めがかからない。薬価抑制や施設基準の見直し等に代表される、大局的な視点から発せられる施策はほぼやりつくされ、悪平等な施策はかえって正当な医療行為の抑制にもつながりかねない。

近年、患者のカルテや診療報酬のレセプトなど、膨大なデータを分析することで、どういった治療をすれば最適な効果が上がるかという「費用対効果」が可視化され、これまでブラックボックスだった『医療のムダ』があぶり出されつつある。スウェーデンでは、高齢社会に備えて社会保障情報も統合し、169もの治療項目をビッグデータで解析し、改革が進められている。

我々は過去の研究により、DPC別の医療コスト分析用データウェアハウスの開発を行い、高速にDPC別医療コストを把握することが可能となった。また我々は、過去の研究で医薬品の相互作用によって引き起こされる有害事象の検知システムも開発した。このシステムによって、データマイニング技法の1つである、アソシエーション分析技法(アプリアリアルゴリズム)を応用し、有害事象が発生した患者の薬歴から、医薬品相互作用の可能性のある薬剤を仮説として提唱することが可能となった。アソシエーション分析技法は、頻出アイテム発見用アルゴリズムとして、マーケットバスケット分析に用いられている。これは消費者の「合わせ買い」の関連性を発見し、オンラインショッピングモールでの商品紹介等で実用化されている。また医薬品の有害事象の発見手法にはベイジアンネットワークも利用されている。

また、近年は人工知能の技術革新が進んでいる。多くは多層化されたニューラルネットワークへの効率的な機械学習を行う「ディープラーニング」法の応用である。

2. 研究の目的

本研究の目的は、電子カルテシステムを含む病院情報システムから発生した、いわゆる「医療ビッグデータ」を格納している病院データウェアハウスから、データマイニング技法の複数のアルゴリズムを応用し、医療コストに影響を与える因子を可視化することで、コストの適正化を支援するシステムを開発することである。

本システムの開発が成功すると、医療の質を担保しつつ医療コストの適正化が実現可能となる。また本システムが普及すると、我が国の医療機関の医療コストの適正化にも寄与するものと考えられる。

3. 研究の方法

本研究では、まず、DPC別医療コスト分析用データウェアハウスシステムを構築した。

本システムには、既設のDPCBANKシステムからDPCBANK情報、院内がん登録システムから、がん登録情報、医事会計システムから会計情報、オーダリングシステムから検査結果情報を抽出し、全て匿名化して格納した。期間は2011年1月~2015年12月までに蓄積されたデータとした。DPCBANKは、鹿児島大学で開発されたDPC別医療コスト分析専用のデータウェアハウスである。

本研究では、特に肝細胞癌の高コスト要因について、クラスタ分析、アソシエーション分析、ベイジアンネットワーク分析、ニューラルネットワーク分析等の手法を適用させ、評価を行った。

4. 研究成果

構築したシステムの性能を評価するために、本研究では肝細胞癌の初回入院・初回治療に関するコスト分析を行った。その理由は、がん治療は概ねプロトコル通りに施行され、コストのばらつきは原則的に発生しにくいと考えられる。したがってコストのばらつきの原因が特定しやすく、かつ、その結果はイレギュラーなものとして、改善の余地が期待できると考えた。

抽出期間は2011年1月~2015年12月までのデータとし、病理診断にてHCCが確定したがん登録由来のデータベースから、入院日、退院日、在院日数、入院時年齢、性別、がんのステージ、手術の有無、化学療法の有無等を情報を抽出した。医療コストデータとしては、がん登録データに記録されていた初回入院、初回治療に該当する入院期間の、費目名と診療点数を費別に抽出した。

データ分析には統計解析ソフトウェア R 64bit 版 ver3.3.0 を用いた。利用したライブラリは、アソシエーション分析(arules ライブラリ)、ベイジアンネットワーク (deal ライブラリ)、ニューラルネットワーク (nnet ライブラリ) である。データ抽出とクレンジングには Perl スクリプトを作成し用いた。

分析用ワークステーションのCPUは、Intel Core i7 6900K@3.2GHz で、主記憶容量は64GByteとした。

肝細胞癌患者数は、365名であった。分析対象となった患者IDの医療コストデータは、約62万行、費目名のバリエーションは3642種であった。

(1) クラスタリングによる分析

まずは、数値データに関するk-means法によるクラスタ分析による特徴群の抽出を試みた。抽出したパラメータは、性別(sex)、在院日数(hospitalization)、がんのステージ(Stage)、年齢(Age)、総医療費(Point)とした。クラスタ分析の結果を図1に示す。

図1によると、明らかに関連性がみられるのは、在院日数(hospitalization)と総医療費(Point)のみであり、がんのステージ(Stage)や年齢(Age)、性別(sex)との有意な

関連性や、特別なクラスは発見されなかった。

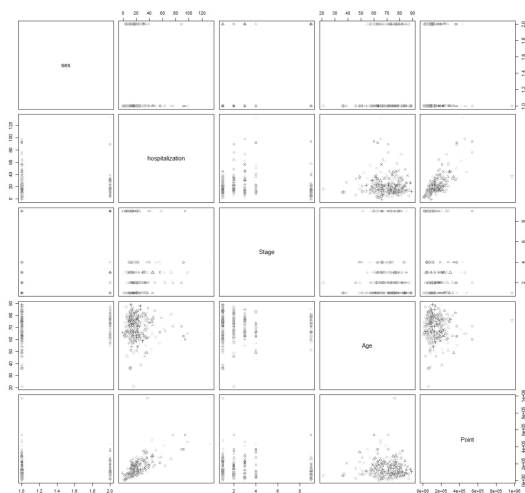


図1 k-means 法による対散布図

この結果を元に、総医療費(Point)を在院日数(hospitalization)で除して、一日当たりの医療費(Point.per.day)を算出し、改めてk-means法のクラスタリングを行った。その結果を図2に示す

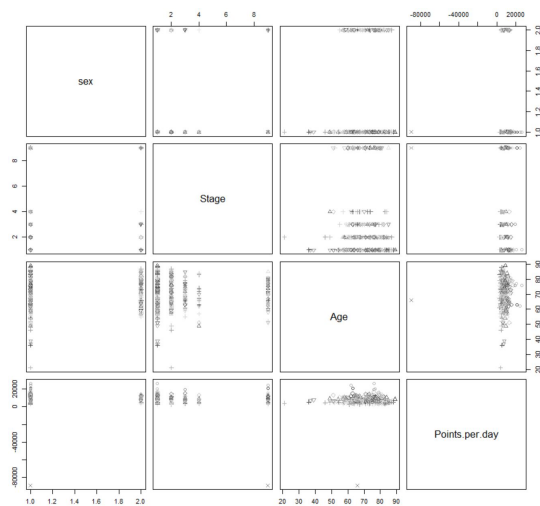


図2 k-means 法による対散布図 その2

この結果からも、明確なクラスは存在せず、一日当たりのコストと関連するパラメータは発見出来なかった。

(2) 高コスト群に特徴的な医療費目の発見に関する分析

クラスタ分析の結果では、明確なクラスは存在しないことが判明したので、1日あたりの医療費が高い群に特徴的な医療費目の発見について、他の複数のデータマイニング手法を用いて分析した。365件のデータの総医療費と在院日数から、一日あたりの医療費を算出し、一日当たりの医療費が大きい順に、高コスト群、中間的コスト群、低コスト群に3分割し、ラベルした。肝細胞癌の患者に行

われた医療行為(薬剤、検査項目を含む)のバリエーションは3642種であった。

まず、アソシエーション分析による高コスト要因分析を試みたが、Item数3642では、メモリアオーバーフローとなり解析不可能であった。分析可能なItem数の上限について調査したところ、96が上限であったので、本目的には利用不可能であると判断した。

次に、ベイジアンネットワークについて要因分析を試みたが、こちらもメモリアオーバーフローにより分析不可能であった。分析可能なItem数は15程度であった。Rは64bit版を採用しており、分析用ワークステーションには64GBByteもの主記憶を準備していたため、これらのメモリアオーバーフローは、ハードウェアの問題ではなく、Rシステムのライブラリの実装上の問題であろうと推測した。

ニューラルネットワーク分析については、Item3642種のままで問題なく分析実行できた。ニューラルネットワークでは、365件のデータを機械学習用データ(トレーニングデータ)274件と、評価用データ91件に均等に分割して分析を行った。

出力結果が高・中・低の3要素なので、中間ニューロン数を4,5と変化させて分析を行った。ニューラルネットワーク分析のイメージを図3(中間ニューロン4)、図4(中間ニューロン5)に示す。なお、このイメージでは、Item数が3642種のままでは作図困難であったため、Item数を10に制限して作図した。

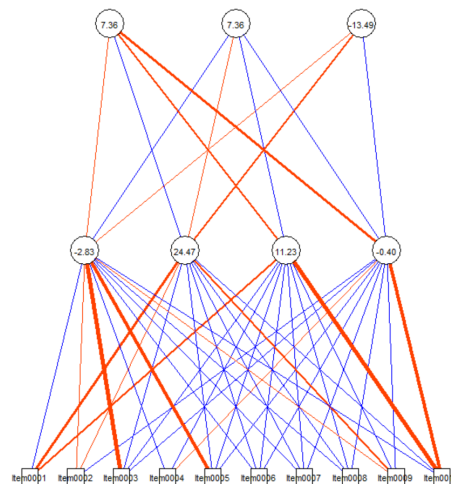


図2 中間ニューロン4の場合の分析例 (Item数を10に限定)

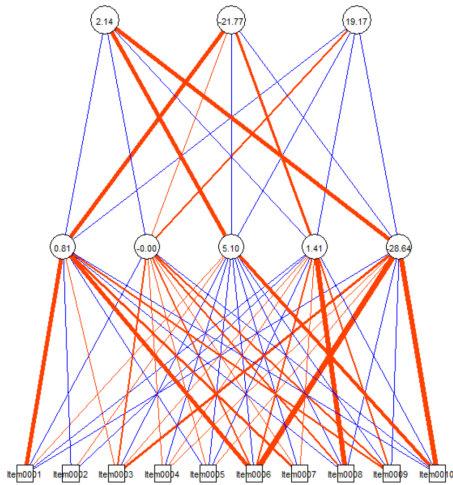


図3 中間ニューロン5の場合の分析例
(Item数を10に限定)

トレーニングデータで学習させたニューラルネットワークモデルで、評価データを評価させた結果を示す。

中間ニューロン数4の分析結果

- ・重み：14587
- ・処理時間：36分19秒
- ・学習試行回数:1090回
- ・予測結果

	High	Mid	Low	sum
High	25	5	0	30
Mid	5	17	9	31
Low	2	11	17	30
sum	32	33	26	91

横軸はニューラルネットワークが費目名のパターンから予測したコスト
縦軸は1日あたりの医療費から算出したコスト

- ・コスト判断一致率 0.6483516

この結果より、MidとLowの判定が上手く処理できていないことが判明したので、Highとそれ以外をまとめて集計した。

	High	Low	sum
High	25	5	30
Low	7	54	61
sum	32	59	91

横軸はニューラルネットワークが費目名のパターンから予測したコスト
縦軸は1日あたりの医療費から算出したコスト

- ・高コスト判定のみの一致率=0.868132

中間ニューロン数5の分析結果

- ・重み：18233
- ・処理時間：1時間14分5秒
- ・学習試行回数：1340
- ・予測結果

	High	Mid	Low	sum
High	25	5	0	30
Mid	9	11	11	31
Low	1	7	22	30
sum	35	23	33	91

横軸はニューラルネットワークが費目名のパターンから予測したコスト

縦軸は1日あたりの医療費から算出したコスト

- ・コスト判断一致率=0.6373626

Highとそれ以外をまとめて集計した結果

	High	Low	sum
High	25	5	30
Low	10	51	61
sum	35	56	91

横軸はニューラルネットワークが費目名のパターンから予測したコスト

縦軸は1日あたりの医療費から算出したコスト

- ・高コスト判定のみの一致率=0.835165

このように、機械学習したモデルを用いて、コスト費目から総医療費が高額になるか否かの判定は80%以上の的中率で可能であることが判明した。中間ニューロンが増えると、処理時間が倍増するが、精度は上がらなかった。

次に、機械学習したニューラルネットワークの入力ニューロンの重みから、高コスト要因となったコスト費目を抽出した。学習した重みの抽出には成功したが、中間ニューロン4、5共に、試行するたびにニューロンの重みは変化していた。一致するItem割合は10%程度であった。このことより、ニューラルネットワークによる機械学習では、高コストに寄与する全ての特徴を捉えているのではなく、高コストになる「断片的な特徴」を捉えていることが分かった。試行毎に発見する特徴が変化する為、機械学習を複数回繰り返し、学習した重みから発見したItemの出現頻度を集計する必要があると思われる。1回の試行に30分以上の時間を要することから、試行の自動化等、更なるシステム開発が必要である。本研究期間では、残念ながらItem頻度を自動計測する部分のシステム開発までは当初の計画にもなく、実現できな

った。今後の研究課題としたい。

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)

[雑誌論文](計6件)

高額医療費の要因分析に適したデータマイニング手法に関する研究: 村永 文学, 岩穴口 孝, 宇都 由美子, 熊本 一郎:医療情報学連合大会論文集 第 37 巻 Suppl. (査読有)(掲載予定)

医薬品相互作用による有害事象シグナル検知システムの分析手法の評価: 村永 文学, 岩穴口 孝, 宇都 由美子, 熊本 一郎:医療情報学連合大会論文集,36 回 1号,446-449(2016.11) (査読有)

DWH の活用における今日的課題と将来展望 鹿児島大学病院の事例を踏まえて;熊本 一郎, 村永 文学;新医療,43 巻 2号,Page24-27(2016.02) (査読有)

病院経営を支える直接・間接生産部門の生産性の評価に関する研究 医療費率から医業収支率による評価導入;宇都 由美子, 岩穴口 孝, 村永 文学, 熊本 一郎:医療情報学連合大会論文集,35 回, 544-545 (2015.11) (査読有)

DPC と医療費率を活用した入院診療体制再評価のための分析; 岩穴口 孝, 宇都 由美子, 村永 文学, 熊本 一郎:医療情報学連合大会論文集,35 回,374-377 (2015.11) (査読有)

The significance of a Bayesian Network in Type 2 Diabetes Mellitus: Nurjannah, Fuminori Muranaga, Takashi Iwaanakuchi, Yumiko Uto, Ichiro Kumamoto; Journal of Joint conference on Medical Informatics, Vol.34 Suppl. Page 352-355 (2014.11) (査読有)

[学会発表](計5件)

高額医療費の要因分析に適したデータマイニング手法に関する研究: 村永 文学, 岩穴口 孝, 宇都 由美子, 熊本 一郎:第 37 回 医療情報学連合大会(大阪府大阪市グランキューブ大阪)にて発表予定(査読有)

医薬品相互作用検知システム用の有害事象シグナル検知アルゴリズムの評価: 村永 文学, 岩穴口 孝, 宇都 由美子, 熊本 一郎:第 36 回 医療情報学連合大会(神奈川県横浜市パシフィコ横浜 2016.11.21) (査読有)

病院経営を支える直接・間接生産部門の生産性の評価に関する研究 医療費率から医業収支率による評価導入;宇都 由美子, 岩穴口 孝, 村永 文学, 熊本 一郎:第 36 回 医療情報学連合大会(神奈川県横浜市パシフィコ横浜 2016.11.21) (査読有)

The significance of a Bayesian Network in Type 2 Diabetes Mellitus: Nurjannah, Fuminori Muranaga, Takashi Iwaanakuchi, Yumiko Uto, Ichiro Kumamoto; The 34th Joint conference on Medical Informatics (千葉県千葉市幕張メッセ 2014.11.6) (査読有)

診療記録の記載事項より誤嚥性肺炎の発症を予見する取り組みとその評価 データマイニングを用いて: 小牧 祥太郎, 村永 文学, 川原 浩, 山下 美穂, 岩穴口 孝, 宇都 由美子, 熊本 一郎, 福田 ゆかり:第 34 回医療情報学連合大会(千葉県千葉市幕張メッセ 2014.11.6) (査読有)

[図書](計0件)

[産業財産権]

出願状況(計0件)

取得状況(計0件)

[その他]

特記事項無し

6. 研究組織

(1)研究代表者

熊本 一郎 (KUMAMORO Ichiro)
鹿児島大学・医歯学域医学系・教授
研究者番号:40225230

(2)研究分担者

村永 文学 (MURANAGA Fuminori)
鹿児島大学・医歯学域附属病院・講師
研究者番号:00325812

宇都 由美子 (UTO Yumiko)
鹿児島大学・医歯学域医学系・准教授
研究者番号:50223582

岩穴口 孝 (IWAANAKUCHI Takashi)
鹿児島大学・医歯学域附属病院・助教
研究者番号:80619198