

一般口演

一般口演21

医療データ分析9（機械学習・テキストマイニング2）

2018年11月25日(日) 10:40～12:10 D会場 (4F 413+414)

[4-D-2-3] 消化器外科手術患者における術後感染の早期検知をめざした機械学習と統計モデルの検討

○中澤 香子¹, 福島 慶子³, 浅倉 宏至³, 齋藤 翔太⁴, 石川 卓², 外山 聡¹, 赤澤 宏平² (1.新潟大学医歯学総合病院薬剤部, 2.新潟大学医歯学総合病院医療情報部, 3.NECソリューションイノベータ, 4.新潟医療福祉大学)

背景：外科治療では、感染予防を目的としてガイドラインに基づき術中・術後に抗菌薬を使用するが、その後感染を起こす患者がいる。本研究では、術後感染発生リスク因子の探索と発症予測精度の評価を行うため、消化器・一般外科の手術患者の病院情報システムに蓄積された診療データを機械学習ツールと統計学的手法で解析した。方法：1) 患者情報の取得；2013年6月1日～2017年5月31日に消化器外科で手術した患者の基本情報および臨床所見、薬剤所見、手術所見を病院情報システムより取得した。2) 取得データのクレンジング；病名はICD-10のアルファベット毎、薬剤データは厚労省コードの上位2桁で分類した。3) 解析；感染的中率及びリスク因子は、NECの機械学習ツール及びステップワイズ法によるロジスティック回帰分析にて解析した。なお、手術日以降の細菌検査結果が陽性の場合感染とした。結果：手術日以前の患者情報と手術中データ、感染陽性：陰性＝1：1とした機械学習ツールでの解析におけるリスク因子は、手術中呼吸器官用薬剤数が1.5以上、代謝性医薬品血液・体液用薬数が37.5以上であり、感染的中率は81.8%だった。患者基本情報、ドレーンの有無、手術時間を用いたロジスティック回帰分析のリスク因子は、手術時間6時間以上、入院期間中の呼吸疾患の病名であった。手術時間1時間未満と比較したオッズ比は、6.06[95% CI；2.78-9.20]、感染発生率は1.62倍、入院期間中に呼吸疾患の病名「あり」群は「なし」群と比較して、オッズ比4.63[95% CI;2.89-7.43]、感染発生率は1.53倍高く、感染的中率は77.4%であった。考察：機械学習ツールでは、手術中データが感染的中率に影響を及ぼした。また、解析ごとにリスク因子が異なった。感染症発生は、早期に検知することが重要となる。今後は病院情報システムの導入へ向けて手術前に得られる患者情報を基に、より精度の高いモデルを構築する。

消化器外科手術患者における術後感染の早期検知をめざした 機械学習と統計モデルの検討

中澤香子^{*1}、福島慶子^{*2}、
浅倉宏至^{*2}、齋藤翔太^{*3}、石川卓^{*4}、外山聡^{*1}、赤澤宏平^{*4}

*1 新潟大学医歯学総合病院薬剤部、*2 NEC ソリューションイノベーター、
*3 新潟医療福祉大学医療情報管理学科、*4 新潟大学医歯学総合病院医療情報部

Development of machine learning models and statistics models to estimate infection positive for surgery patients

Kyoko Nakazawa^{*1}, Keiko Fukushima^{*2}, Kouji Asakura^{*2}, Shota Saito^{*3}, Akira Toyama^{*1}, Kohei Akazawa^{*4},

*1 Department of Pharmacy, Niigata University Medical and Dental Hospital, *2, NEC Solution Innovators

*3 Department of Health Informatics, Niigata University of Health and Welfare *4 Department of Medical Informatics,
Niigata University Medical and Dental Hospital

Abstract in English comes here.

[Backgrounds] The anti-bio drugs was used surgery patients to prevent surgical site infection. But the patients often occur it. Therefore The diagnosis of surgical site infection is difficult. This study aimed to estimate infection positive patient to develop machine learning models, and compared with statistics models.

[Methods] The data from 1,789 patients, who performed surgery were analyzed in this study. We developed machine learning models which use patient demographics, diagnoses, mediations and laboratory tests to classify the patients into 2 categories whether they have a infection or not. And we constructed by logistic regression analysis.

[Result] The machine learning models gained 81.8% of infection positive rate. Logistic regression analysis was obtained 77.4%, The odds rate of the set of information for patients, operate time and to diagnose infection positive was 6.06 (95% CI: 2.78-9.20).

[Discussion] The machine learning models effect on drugs that use during patients received surgery. Logistic regression analysis was effect on long time of surgery.

Keywords: Machine learning, Logistic regression analysis, Infection.

1. 背景

手術患者における感染症発生を予防するため、ほぼすべての症例において抗菌薬を使用する。抗菌薬は適切な種類、投与日数で投与する必要がある。その理由として、感染症の悪化や耐性菌の発生予防、入院の長期化、薬剤費の軽減が挙げられる。従って手術時から抗菌薬の投与、創部の清潔管理、ドレーン管理を行い、感染症を予防する治療やケアが求められる。しかしそのような管理を実施しているにもかかわらず、感染症が発生する場合がある。

術後感染予防抗菌薬適正使用のための実践ガイドライン[1]では、手術目的別に推奨抗菌薬や投与量、投与日数が掲載されている。しかし、胆嚢炎や肝臓瘍など手術前より感染を持っている場合や高齢者、ステロイド使用者などの術後感染高リスク患者においては、使用抗菌薬を検討する必要がある。現在これらの判断は、臨床検査や患者の容態に応じた診断や医師の経験が重要な部分を占める。

そこで、病院情報システムに蓄積された情報を基に、術後感染症の早期に予測できるような機械学習及び統計学手法を開発することを試みた。

2. 目的

本研究では、術後感染発生リスク因子の探索と発症予測精度の評価を行うため、消化器・一般外科の手術患者の病院情報システムに蓄積された診療データを機械学習ツールと統計学的手法で解析した。

3. 方法

3.1. 患者情報の取得

2013年6月1日から2017年5月31日までの4年間に消化器外科で手術を実施した患者のデータを病院情報システムより抽出した。該当患者は1789名であった。患者情報は、入院中の、検査値、薬剤の種類、投与量、日数、手術時間、細菌検査の材料や培養結果とした(表1)。

表1 病院情報システムから取得した情報(一部抜粋)

患者基本情報	手術時情報	入院期間中情報
年齢	手術日	病名
血液型	手術中使用薬剤	体温
身長	手術時間	ドレーンの有無
体重	術式	臨床検査結果
入院日		細菌検査
退院日		
転帰		

3. 2. データの構成

取得したデータを解析可能なデータ一覧となるよう整理した。

必要に応じて変数をカテゴリ化(例;基準値より低い、基準値内、基準値より高い、未設定)した。病院情報システム内でカテゴリ化されているものはそのまま利用した。臨床検査値の基準値は、日本で一般的に用いられる日本臨床検査標準協議会(JCCLS)とした。クレアチニンクリアランスは Cockcroft-Gault 式を用い、女性は 0.8 乗じた。

病院情報システムに自動入力されない項目(身長や体重)は、外れ値がないか前後の入力データと比較し確認した。外れ値があった場合は、病院情報システムで確認し、誤入力の場合は、除外した。身長の入力データがない場合は、厚生労働省より各年齢の平均値を代用した。

病名は ICD コードの大分類毎に分類した。

薬剤データ(処方オーダー、注射オーダー)は、薬価収載されている医薬品に付随する厚労省コードの上位 2 桁でカテゴリ化した。例えば、ロキソニン錠 60mg の厚労省コードは「1149019F1560」であるが、この場合は「11」に分類した。厚労省コードは薬効分類(頭4桁)で付与されているため、類似医薬品を分類することが可能である。

なお、手術時に使用した薬剤は、医事データより抽出した。薬剤の選定は、該当期間の手術室定数配置薬に指定されていたものとした。

ドレーン挿入期間は、2017 年 1 月以降は処置オーダーの記録を用いた。それ以前は医事データでドレーンの処置に対して算定した日をドレーン挿入期間とした。

3. 3. 解析

機械学習ツール(3.4 参照)と統計学的手法で解析した。統計学的な解析は、単変量解析は R を、多変量解析は SPSS を用いた。

手術日以降、細菌検査の結果が陽性となった患者を『感染あり』とした。感染日は細菌検査の検体採取日とした。

3. 4. 利用した機械学習ツール

本研究では、異種混合学習技術[2]を用いて異種混合予測モデルを探索した。異種混合予測モデルとは、入力データを決定木式のルールによって場合分けし、各場合で異なる説明変数を組み合わせた線形モデルで予測するモデルである。異なる説明変数の組み合わせ(異種)による予測モデルを組み合わせて(混合)予測する。このモデルでは、予測式の組み合わせ、説明変数の組み合わせ、データの場合分けを行う。

異種混合学習技術とは、因子化情報基準と因子化漸近ベイズ推論による異種混合モデルの学習技術の総称である。

因子化情報基準は、予測モデルの良さを判定する。一般的には、赤池情報基準、ベイズ情報基準が用いられるが、これらは単一のモデルで予測する。異種混合モデルは、複数のモデルを切り替えながら予測することから、独自の情報基準を導出している。因子化情報基準は、予測精度が低い、あるいはデータの場合分けや説明変数の組み合わせが複雑なモデルは値が低くなり、シンプルで予測精度の高いモデルは値が高くなる特徴を持つ。

また、因子化情報基準を最大化するモデルを探索するアルゴリズムは、因子化漸近アルゴリズムを使用する。これは、①データの場合分け条件の最適化、②説明変数の組み合わ

せの最適化、③不要な予測モデルの削除の3ステップを繰り返すことで因子化情報基準を最大にするモデルを探索するアルゴリズムである。このアルゴリズムを用いることで、各ステップで必ず因子化情報基準の値を改善し、高速に最適な異種混合予測モデルの発見が可能となる。

4. 結果

4. 1. 基本分析

DPC 病名の登録のあった 1788 名を対象とした。細菌検査に基づきデータを作成した結果 2078 件を解析対象とした。そのうち『感染あり』の患者は 661 件(31.8%)、『感染なし』が 1417 件(68.2%)であった。

感染ありとなったタイミングは入院時 82 件、入院後術前 214 件、術後感染 365 件であった。

それぞれの項目の有意差検定(u 検定)を実施したところ、年齢、手術時間、ドレーンの有無、手術時薬剤(代謝性医薬品_血液・体液用薬)の使用は有意な差があった。

表 2 各項目の検定結果

項目	P 値
年齢	<0.01
手術時間	<0.01
BMI	0.0107
ドレナージ	<0.01
手術時薬剤(代謝性医薬品 血液・体液用薬)の使用	<0.01

4. 2. 機械学習ツールを用いた結果

手術日以前の患者情報と手術中データ、感染陽性:陰性=1:1 とした機械学習ツールでの解析におけるリスク因子は、手術中呼吸器官用薬剤数が 1.5 以上、代謝性医薬品血液・体液用薬数が 37.5 以上であり、感染的中率は 81.8%だった。

4. 3 多変量解析を用いた結果

患者基本情報、ドレーンの有無、手術時間を用いたロジスティック回帰分析のリスク因子は、手術時間 6 時間以上、入院期間中の呼吸疾患の病名であった。手術時間 1 時間未満と比較したオッズ比は、6.06[95%CI;2.78-9.20]、感染発生率は 1.62 倍、入院期間中に呼吸疾患の病名「あり」群は「なし」群と比較して、オッズ比 4.63[95%CI;2.89-7.43]、感染発生率は 1.53 倍高く、感染的中率は 77.4%であった。

5. 考察

現在術後感染症のリスクはガイドラインで定められている。感染症の高リスク条件として、糖尿病、免疫抑制薬やステロイド剤の使用、長時間の手術、肥満、が主要な項目である。しかし実際に感染発生や治療は医師の診断が重要である。

本研究では、術後感染症の有無を予測するモデルを、病院情報システムから得た情報を基に機械学習と統計解析から構築することを試みた。

機械学習ツールでは、手術時間含む手術中データが感染の中率に影響を及ぼすことが判明した。多変量解析では手術時間および入院期間中の呼吸疾患であった。結果より機械学習ツールと統計学的手法では、リスク因子となる変数が異なった。

本研究では、細菌検査の結果の陽性のみで感染の有無を決定した。しかし、起因菌によって使用薬剤が異なること、細菌によってリスクが異なることを考慮し、抗生剤の投与が必要な患者の推定を行う必要がある。また、ガイドライン上の術後感染症の高リスク条件である、糖尿病、免疫抑制薬やステロイド剤の使用について、さらなる検討を行い、さらなる精度の向上を目指す。

また酒井ら[3]は、ニューラルネットワークとロジスティック回帰モデルで虫垂炎発症の予測に関する比較を行っている。報告では、実際の診断とモデルによる診断の一致率において、2値変換した変数を用いた場合、実際の診断がニューラルネットワークのみ、ロジスティック回帰モデルの診断のみ一致したケースに有意差があったとしている。本研究における機械学習ツールとロジスティック回帰分析での実際の診断との一致、不一致についてさらなる検討を重ねることで、より精度が高まる可能性がある。

臨床現場において術後感染の予測の精度が高くかつ早期の検知が重要となることから、手術前に得られる変数でより精度の高いモデルの構築を行う必要がある。

6. 倫理配慮・謝辞

本研究は新潟大学医学部倫理委員会の承認を得て実施した。

新潟大学医歯学総合病院感染管理部には本研究のデータ解析に関してご指導を頂きました。

参考文献

- 1) 「術後感染予防抗菌薬適正使用のための実践ガイドライン」
公益社団法人日本化学療法学会/一般社団法人日本外科感染症学会
- 2) 藤巻遼平、森永聡、江藤力. 異種混合学習技術とビッグデータ分析ソリューションの研究開発
- 3) Sakai S, Kobayashi K, Toyabe S, Mandai N, Kanda T, Akazawa K. Comparison of the levels of accuracy of an artificial neural network model and a logistic regression model for the diagnosis of acute appendicitis. J Med Syst 2007; 31(5): 357-364.

