

---

一般口演

## 一般口演26

### 機械学習・HISデータ解析

2017年11月23日(木) 09:15 ~ 10:45 F会場 (10F 会議室1004-1005)

---

#### [4-F-1-OP26-1] 機械学習を用いた感染症予測モデルの構築と検証

田中 龍也<sup>1,2</sup>, 山下 哲平<sup>1</sup>, 田中 伸<sup>1</sup>, 近藤 学<sup>3</sup>, 佐藤 仁<sup>3</sup>, 池田 睦<sup>3</sup> (1.滋慶医療科学大学院大学, 2.社会医療法人 祐生会 みどりヶ丘病院, 3.NECネクサソリューションズ株式会社)

##### 目的：

病院における感染防止対策は、患者が安全に医療を受けるために必要な業務である。しかしながら、看護師や医師が感染の疑いを判断し注意を払うことは、日々の業務量から考えても現実的ではない。そのため、過去の電子カルテ情報を用いて、各患者の感染症の発症確率の高さを予測することにより、感染症の判断を補佐するシステムを実現できないかと考えた。そこで本研究では、国際統計分類 ICD-10における A09「感染症と推定される下痢及び胃腸炎」を対象とし、電子カルテから出力された SS-MIX II データで取得できる情報を元に A09を判別するモデルの構築とその検証を行った。

##### 方法：

2011/4/18から2016/6/22までの期間に入院した患者194,100名のうち、ランダムに抽出した155,643名(学習群：A09罹患患者4,411名)を判別モデルの学習に用い、残りの44,467名(検証群：A09罹患患者1,012名)により判別モデルの性能を検証した。判別に用いる入力情報としては、匿名化した患者基本情報、外来時と入院時の過去の傷病履歴（A09以外）などを用いた。機械学習および判別を行うソフトウェアとして、NECのRAPID機械学習を使用し、NECネクサソリューションズ株式会社と実施した。

##### 結果：

検証群を用いて、入力変数から患者のA09への感染確率を計算し、0.5以上であれば陽性と判別した。A09感染者1,012名のうち陽性と判別されたものは804名（感度79.4%）、非感染者43,455名のうち陽性と判別されたものは5,893名（偽陽性率13.6%）であった。また、偽陽性率が5%となるカットオフ値( $P=0.899$ )を用いたとき、感度は52.3%であった。

##### 考察：

感染症の確率を計算することにより、日常の感染管理をサポートできる可能性が示された。今後は、薬剤の履歴など他の情報を加えることにより、さらに精度を高め、確率に応じた患者隔離などの感染管理を実施した場合のコストなどを検討していく。

# 機械学習を用いた感染症予測モデルの構築と検証

田中 龍也<sup>\*1 \*2</sup>、山下 哲平<sup>\*1</sup>、田中 伸<sup>\*1</sup>、  
近藤 学<sup>\*3</sup>、佐藤 仁<sup>\*3</sup>、池田 睦<sup>\*3</sup>

\*1 滋慶医療科学大学院大学、\*2 社会医療法人 祐生会 みどりヶ丘病院、  
\*3 NEC ネクサソリューションズ株式会社

## Using machine learning to construct and verify an infectious disease prediction model

Tatsuya Tanaka <sup>\*1 \*2</sup>, Teppei Yamashita <sup>\*1</sup>, Shin Tanaka <sup>\*1</sup>,  
Manabu Kondo <sup>\*3</sup>, Hitoshi Sato <sup>\*3</sup>, Mutsumi Ikeda <sup>\*3</sup>

\*1 Graduate School of Health Care Sciences, Jikei Institute,  
\*2 Midorigaoka Hospital, \*3 NEC Nexsolutions, Ltd.

**Aim:** Hospitals have to implement infection prevention measures so that patients can be cared for safely. However, it is unrealistic to expect nurses and doctors to be constantly on their guard against possible sources of infection in addition to their daily workload. We therefore considered the possibility of implementing a system that can assist in the judgment of infectious diseases by predicting the probability of infectious diseases developing in each patient based on previous electronic medical chart information. In this study, we focused on code A09 (Diarrhea and gastroenteritis of presumed infectious origin) of the International Statistical Classification of Diseases and Related Health Problems (ICD-10), and we constructed and verified a model for judging A09 based on information obtainable from SS-MIX II data output from electronic medical charts.

**Method:** From the electronic medical chart information of 219,253 patients who were admitted to hospital during the period from April 18, 2011 to June 22, 2016, we randomly selected 178,535 cases to train the discrimination model (training group: 3,749 A09 patients), and we used the remaining 44,467 cases (verification group: 1,012 A09 patients) to verify the performance of the discrimination model. The input data used for discrimination included anonymized basic patient details and medical records of previous visits to the hospital as in-patients and out-patients (other than A09). Machine learning and discrimination was performed together with NEC Nexsolutions using NEC's RAPID machine learning software.

**Results:** Using the verification group, we calculated the A09 infection probability of patients based on the input data, and recorded a positive match in cases with a probability of 0.5 or more. Out of the 1,012 cases of A09 infection, 804 were judged to be positive matches (sensitivity 79.4%), and out of 43,455 non-infected cases, 5,893 were judged to be positive (false positive rate 13.6%). Also, when using a cut-off value for which the false positive rate was 5% ( $P=0.899$ ), the sensitivity was 52.3%.

**Discussion:** We have shown that it is possible to support routine infection control by calculating the probability of infectious disease. In the future, we aim to achieve greater precision by adding additional information such as medication history, and we will investigate the costs of implementing infection control measures such as isolating patients according to the probability of infection.

**Keywords:** Machine learning, A09, Infectious disease, SS-MIX II.

### 1. 目的

病院内における感染症の防止対策は、患者が安全に医療を受けるために必要な業務である。そのため、感染制御の方向性を示す厚生労働省の通知が幾度と行われ、診療報酬において感染防止対策加算の施設基準が示されるなど、感染を防止するための試みが数多く為されてきた<sup>1)</sup>。しかしながら、近年においてもインフルエンザや薬剤耐性腸内細菌などのアウトブレイクの発生について数多く報告されている<sup>2) 3) 4)</sup>。

ひとたび院内においてアウトブレイクが発生してしまうと、感染患者へのケアが必要になるのは当然として、対応する人員の確保により通常の業務へ支障が生じること、病棟の閉鎖などによる経営上の損失が増大すること、社会的信用を喪失す

ることなど、さまざまな損失を病院に与え、経営を継続することに対して重大な支障を生じることとなる。

感染防止のためには、まず感染した患者を発見し隔離することが重要であるが、看護師や医師が感染の疑いを判断し注意を払うことは、日々の業務量から考えても現実的ではない。通常の業務に影響を与えずに、自動的に感染者を発見できるシステムが必要となる。

そのため、過去の電子カルテ情報を用いて、各患者の感染症の発症確率の高さを予測することにより、感染症の判断を補佐するシステムを実現できないかと考えた。電子カルテに入力済みの情報から自動的に予測するのであれば、新しい業務を追加する必要もないため、導入も容易である。

そこで、本研究では国際統計分類 ICD-10 における A09 「感染症と推定される下痢及び胃腸炎」を対象として、電子カルテから出力された SS-MIX II データで取得できる情報を元に A09 を判別するモデルの構築とその検証を行った。

## 2. 方法

2011/4/18 から 2016/6/22 までの期間に診察した患者を対象として、A09 感染判別モデル(以下、判別モデル)の学習と検証を行った。まず、電子カルテ DB より SS-MIX 標準化ストレージより、219,253 名のデータを HL7 形式で抽出し、分析用の環境に移動した。

219,253 名のうち、ランダムに抽出した 174,786 名(学習データ)を判別モデルの学習に用い、残りの 44,467 名(検証データ)により判別モデルの性能を検証した(図 1)。

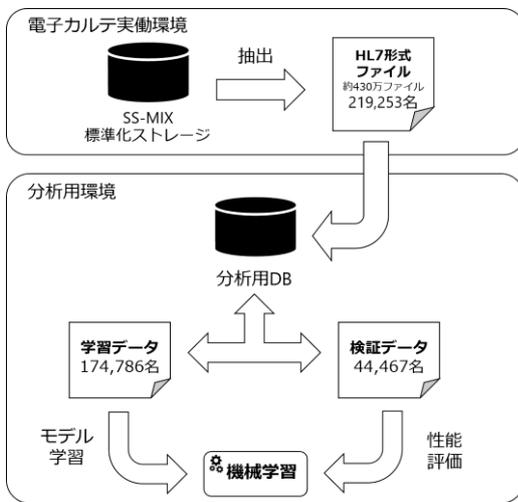


図 1 分析データの流れ

判別に用いる入力情報としては、SS-MIX 標準化ストレージに含まれる情報を用いるということで、患者の基本情報として、生年、性別、近親者続柄、加入保険プラン、過去の傷病履歴として外来時と入院時の過去の A09 以外の疾患の情報を用いた。

機械学習および判別を行うソフトウェアとして、NEC の RAPID<sup>5)</sup> 機械学習を使用し、分析は NEC ネクサソリューションズ株式会社と実施した。

判別モデルは、5 層(入力層、隠れ層 1~3、出力層)のディープニューラルネットワーク(図 2)を用いた。

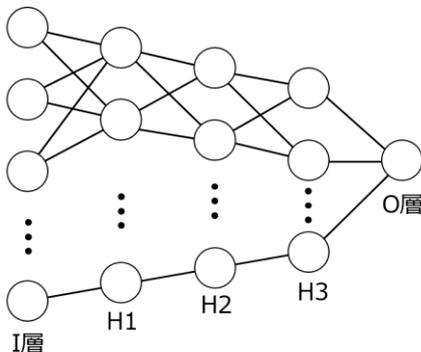


図 2 判別モデル

入力層(I 層)のノード数は、生年と性別(2 変数)、近親者続柄(10 変数)、加入保険プラン(22 変数)、外来時と入院時の傷病履歴(846 変数×2)の 1,726 である。また、隠れ層のノード数は、それぞれ H1 が 1,294(入力層の 3/4)、H2 が 863(入力層の 2/4)、H3 が 431(入力層の 1/4)とし、ノード間の結合は全結合とした。

## 3. 結果

学習データの 174,786 名のうち A09 罹患者は 3,749 名(2.1%)であり、検証データの 44,467 名のうち A09 罹患者は 1,012 名(2.3%)であった。

学習データを用いて判別モデルを学習させたのち、検証データを用いて、入力変数から各患者の A09 への感染確率を予測した。感染確率による判別性能を確認するために作成した ROC 曲線と感度特異度曲線を図 3 に示す。ROC 曲線の曲線化面積は 0.89 であった。

感染確率が 0.5 以上であれば陽性と判別した場合、A09 感染者 1,012 名のうち陽性と判別されたものは 804 名(感度 79.4%)、非感染者 43,455 名のうち陽性と判別されたものは 5,893 名(偽陽性率 13.6%)であった。また、偽陽性率が 5%となるカットオフ値(P=0.899)を用いたとき、感度は 52.3%であった。

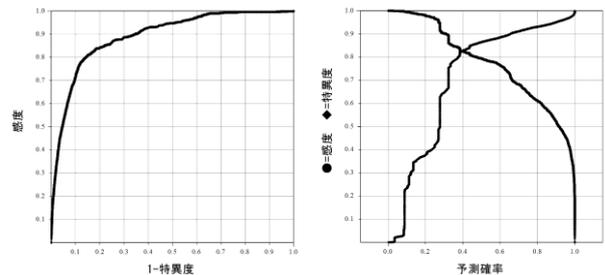


図 3 検証データの ROC 曲線と感度特異度曲線

## 4. 考察

感染症の確率を計算することにより、日常の感染管理をサポートできる可能性が示された。今後は、薬剤の履歴など他の情報を加えることにより、さらに予測精度を改善し、感染確率の予測に応じた患者隔離などの感染管理を実施した場合のコストなどを検討していく。

## 参考文献

- 1) 大久保憲. わが国の感染制御関連施策の変遷とその背景. 日本環境感染学会誌 2016; 31: 213-23.
- 2) 井上三四郎. 急性期整形外科病棟で生じたインフルエンザアウトブレイク. 整形外科と災害外科 2017; 66: 312-8.
- 3) 金城真一. 大学病院でのインフルエンザによるアウトブレイクの経験. 滋賀医大誌 2014; 27: 14-8.
- 4) 矢内充. 院内感染(医療関連感染). 日大医誌 2017; 76(3): 121-4.
- 5) NEC Advanced Analytics - RAPID 機械学習, <http://jpn.nec.com/rapid/>