

ポスター | 知識工学／医療データ解析

ポスター1

知識工学／医療データ解析

2019年11月22日(金) 14:50 ～ 16:20 ポスター会場2 (国際展示場 展示ホール8)

[2-P2-1-03] 電子カルテのデータベースと外部データを組み合わせたインフルエンザ検査キット需要数予測の試み

○吉岡 克己¹、石渡 葉子¹、西周 裕晃¹、伊坂 大樹¹、小笠原 明¹、野本 和宏¹、近藤 学²、武田 一学²、種井 優次²
(1. 公立長生病院, 2. NECネクサソリューションズ株式会社)

キーワード : AI machine learning, SS-MIX Standard Storage, external data, influenza test kit

当院（公立長生病院）は、平成27年度に NECの電子カルテ「MegaOak-MI・RA・Is」を導入し、平成28年2月より本稼働している。以来、電子カルテから出力された SS-MIX標準化ストレージに様々なデータが蓄積されている。このデータを AI処理することにより、インフルエンザキットの需要数を予測できるのではないかと考え、研究を行ったので、報告する。

院内の SS-MIX標準化ストレージに蓄積された院内データに加え、インフルエンザの患者数に影響を及ぼすであろうと予測され、かつ入手が容易である、気象庁より公開されている最高・最低気温、湿度、降水量などの外部データを AI解析の要素として、用いた。過去のデータを学習用と評価用のデータに分け、分析エンジンで学習し、予測モデルを作成した。次に評価用のデータに対し、予測モデルを用いて予測を実行し、実際の検査数と比較検討を行った。なお、データの分析エンジンは、NECの AIソフトウェア製品群の一つである RAPID 機械学習¹⁾を利用した。

その結果、電子カルテのデータベースと外部データを組み合わせることにより、インフルエンザキットの需要予測を行うことができる可能性が高いことが示唆された。

1) NEC Advanced Analytics - RAPID 機械学習,

<https://jpn.nec.com/rapid/>

電子カルテのデータベースと外部データを組み合わせたインフルエンザ検査キット の需要数予測の試み

吉岡克己^{*1}、石渡葉子^{*1}、西周裕晃^{*1}、伊坂大樹^{*1}、野本和宏^{*1}、小笠原明^{*1}、
武田一学^{*2}、武田恵^{*2}、種井優次^{*2}、近藤学^{*2}

^{*1} 公立長生病院、^{*2} NEC ネクサソリューションズ株式会社

Predicting the required amount of influenza test kit using electronic medical record database and external data

Katsumi Yoshioka^{*1}, Yoko Ishiwatari^{*1}, Hiroaki Saisyu^{*1}, Hiroki Isaka^{*1}, Kazuhiro Nomoto^{*1}, Akira Ogasawara^{*1},
Kazunori Takeda^{*2}, Megumi Takeda^{*2}, Yuji Tanei^{*2}, Manabu Kondo^{*2}

^{*1}Chosei Municipal Hospital, ^{*2} NEC Nexsolutions, Ltd.

Abstract

We predicted the required number of influenza test kits by AI analysis. The data used for analysis is an anonymized electronic medical records database (SS-MIX Standardized Storage) and external data published by the Japan Meteorological Agency.

The analysis used "RAPID machine learning", one of NEC's AI software products.

Comparing the predicted value with the actually measured value, 75% of the total had an error rate of less than 30%.

There are 30 influenza test kits in one set. It was confirmed that practical prediction was possible by combining not only data obtained from electronic medical records but also external data.

Keywords: AI machine learning, SS-MIX Standardized Storage, external data, influenza test kit

1. はじめに

当院(公立長生病院)は、2015 年度に NEC の電子カルテ「MegaOak-MI・RA・Is」を導入し、2016 年 2 月より本稼働している。以来、電子カルテから出力された SS-MIX 標準化ストレージに様々なデータが蓄積されている。このデータを広く一般に公開されている様々なデータと組み合わせて AI 処理することにより、インフルエンザ検査キットの需要数を予測できるのではないかと考え、研究を行ったので、報告する。

2. 方法

2.1 データの収集

本研究に用いたデータは、以下の 4 つである。

- ① 診療情報: 過去の診療情報が保持するインフルエンザ検査キットの使用状況を匿名化処理した SS-MIX 標準化ストレージより抽出した。データの抽出期間は、電子カルテ本稼働前の 2015 年 12 月から本稼働中の 2018 年 4 月までである。
- ② 発注情報: 過去の発注情報からインフルエンザ検査キットの発注実績を抽出した。発注データの抽出期間は 2016 年 4 月から 2018 年 4 月までである。
- ③ 気象情報: 過去の診療情報と合致する期間の気象情報(最高気温、最低気温、最高湿度、最低湿度、降水量、日照時間など)を気象庁のホームページ¹⁾から入手した。
- ④ 定点あたりのインフルエンザ報告者情報: 過去の診療情報と合致する期間のインフルエンザ流行状況を把握する指標として、千葉県感染症情報センター²⁾が公開している定点あたりのインフルエンザ報告者数も参照した。

以上の 4 つのデータのうち、①と②は、院内データで、③と④は外部のオープンデータである。

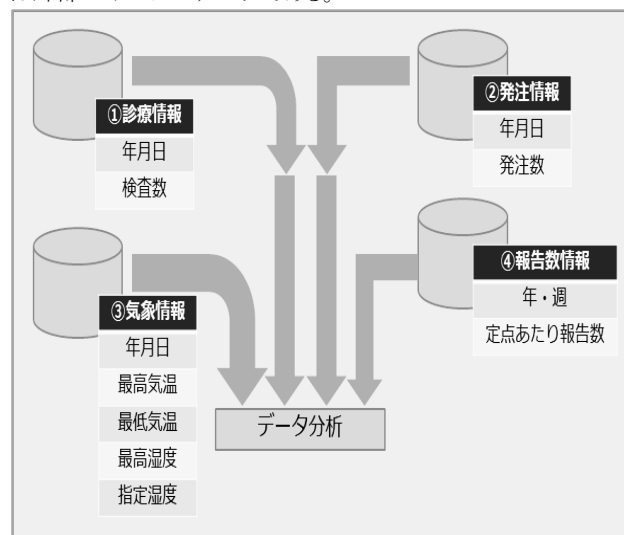


図1 予測に用いる要素の検討

2.2 データ分析の方法

4 つのデータの各項目とインフルエンザ検査数を比較し、傾向分析を行い、関連を検討した(図1)。

この項目の中から有用と思われる項目を選び、過去のデータを学習用と評価用のデータに分け、学習用のデータによって学習を行い、予測用データの分析モデルを生成した。予測値と実測値を比較した値(平均誤差)が小さくなるように分析エン

ジンで学習を繰り返し、平均誤差の値が収束していると判断できた時点で学習を終了し、予測モデルを作成した。

次に評価用のデータに対し、予測モデルを使って予測を実行し、実際のインフルエンザ検査数と比較検討を行った。なお、データの分析エンジンは、NECのAIソフトウェア製品群の一つであるRAPID 機械学習³⁾を利用した。

3. 結果と考察

3.1 傾向分析

図2に過去のインフルエンザ検査キットの発注実績とインフルエンザの検査数の素データを示す。検査数は、当然ではあるが、インフルエンザの流行期である秋から冬にかけて多くなり、12月末から2月にかけてピークを迎える。それに対し、発注数は必ずしもピークと一致していない。発注は、必要な時にインフルエンザ検査キットが足りなくなることがないようにあらかじめ、多く発注していることと、随時発注するのではなく、まとめて発注することの影響が考えられる。すなわち、発注や在庫管理が合理的に行われていないとも言える結果である。

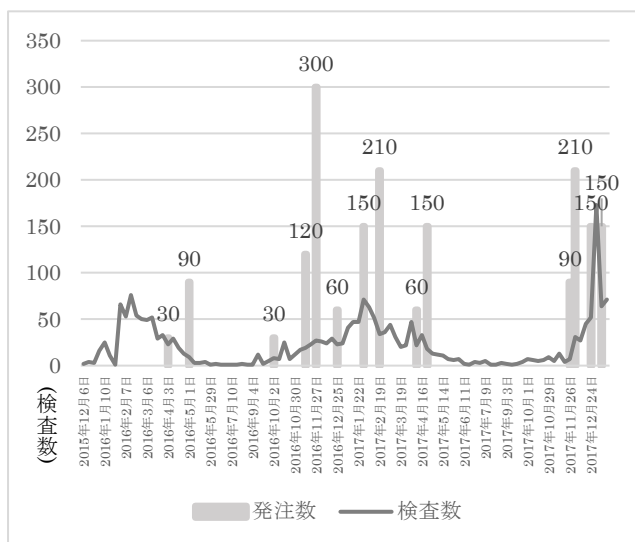


図2 インフルエンザの発注数と検査数

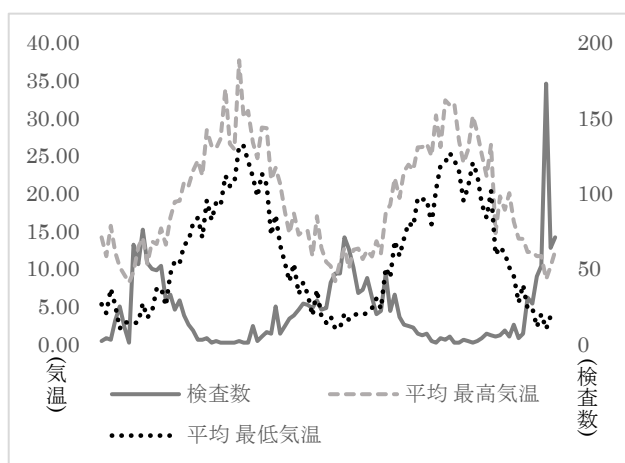


図3 気温とインフルエンザの検査数

図3、4のごとく、最高気温・最低気温・平均湿度の項目とインフルエンザ検査数は、上昇とともに検査数が減少し、下降とともに検査数が増加する、負の相関関係の傾向を認めた。

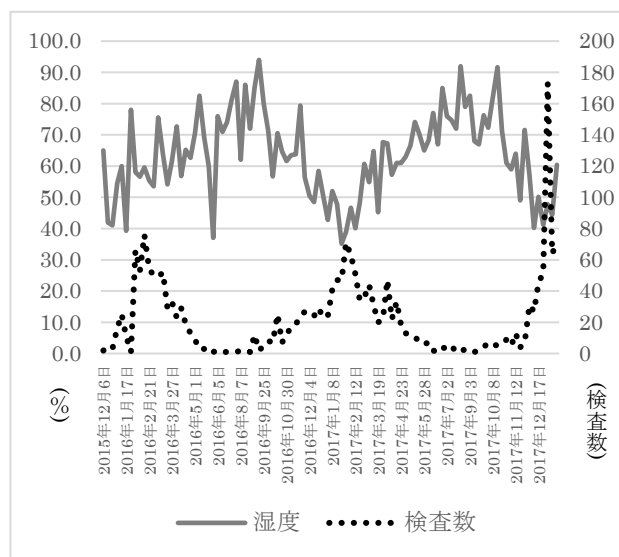


図4 湿度とインフルエンザの検査数

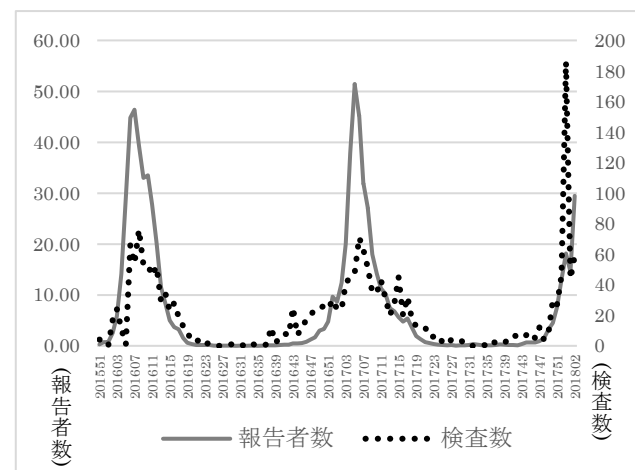


図5 定点あたり報告者数と当院での検査数

図5は、千葉県感染症情報センターが公開しているインフルエンザの定点あたりのインフルエンザ報告者数と当院での検査数との比較である。横軸の数字は、上4桁が西暦の年を、下2桁が第何週目かを表している。縦軸の数字は、左側が定点あたりの報告者数、右側が当院での検査数を表している。

図5から読み取れるように、定点あたりのインフルエンザ報告者数と当院での検査数の関係についても、正の相関の傾向が示唆された。なお、定点とは、インフルエンザ定点のことで、千葉県が指定した千葉県内の215カ所の小児科・内科を標榜する診療所等を指し、定点の医師は、インフルエンザと診断した患者を保健所へ報告することとなっている。したがって、当院でのインフルエンザの流行状況は千葉県内の平均的な流行状況と一致していると考えられる。

3.2 予測試行

今回、需要予測を行うにあたり、これらの傾向分析の結果から、相関が示唆されるものを「説明変数」という予測のための要素として採用した。即ち、最高気温、最低気温、最高湿度、最低湿度といった気象情報、および、定点あたりの報告者数を、過去の検査数実績と合わせて学習用、評価用のデータとして用意した。予測する対象は、当該週の検査数である。

また、より精度を高めるため、過去3年分の要素と、定点あたりの報告者数については、過去17年間の平均からみた平年比についても併せて用意することとした。

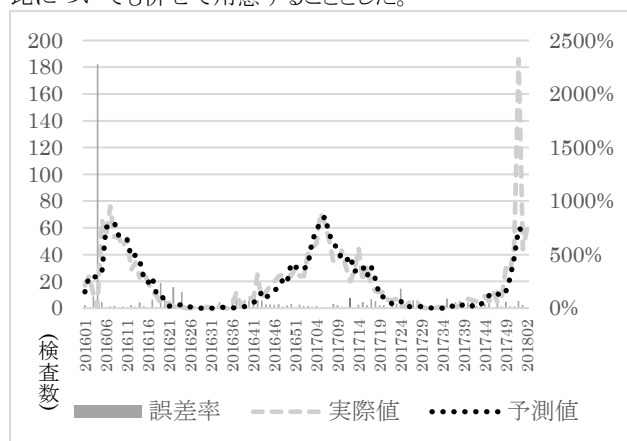


図6 予実差異(学習用データ)

これらの収集データを、学習用データと予測用データとに分けて、分析エンジンにかけて結果を検証した。まず学習用データを使った、予測値と実際の値の比較を図6に示す。グラフからは、予測値は実際値に対し、近似の値で追隨できていることがわかる。

図7は、この学習用データによって生成した分析モデルを利用して、評価用データに対して予測を実施した結果である。予測値と実際値の差異を比較したところ、全体の75%については誤差率が30%未満の結果が得られた。差異が大きなケースも見受けられるものの、インフルエンザ検査キットが1セット30回分であることを考慮すれば、現状の環境でも十分実用に耐えられる数値と考えられる。

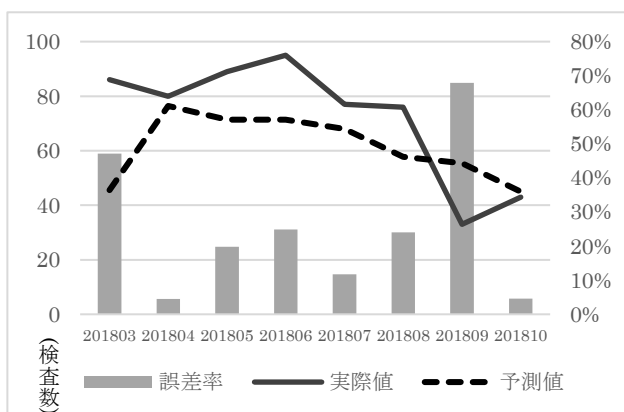


図7 予実差異(予測用データ)

まとめ

以上より、当院に蓄積されている過去の診療情報や、発注情報といった院内データに、気象情報や定点あたりのインフルエンザ報告者情報といった外部のオープンデータを加えることで、インフルエンザ検査キットの適正な発注数を予測できる可能性が高いことがわかった。しかし、本テーマにおいて、より精度の高い結果を求めるための学習および検証を行うには、2年間前後のデータ収集期間では十分なものとはいえない。また、年ごとに異なる「インフルエンザの流行動向」を、予測に投影させる

ためには、十分な期間のデータに加えて、予測を補強するための新たな要素が必要だと考えている。

今後、実用に即した予測を実現していくために、データの収集期間をより長くし、予測精度を高めるうえで有用な新規の要素について、そのデータ収集の可否も含めて検討していくことが今後の課題である。

本テーマにおける成果が、将来的には院内の他の物品の在庫管理にも応用できるような、横への展開がはかれることを期待している。

参考文献

- 1) 気象庁ホームページ
<https://www.jma.go.jp/jma/menu/menureport.html>
- 2) 千葉県感染症情報センターホームページ
<https://www.pref.chiba.lg.jp/eiken/c-idsc/>
- 3) NEC Advanced Analytics - RAPID 機械学習
<https://jpn.nec.com/rapid/>