
ポスター発表

[PB] ポスター B

2020年6月5日(金) 09:00 ~ 18:30 ポスター会場(2) (e-poster)

[PB-19] 関西地域 86 薬局における抗インフルエンザ薬処方件数の予測 Prediction of Prescription Counts of Anti-influenza Drugs on 86 Pharmacies in Kansai Area

*白方 大樹¹、西山 悠¹、藤林 和俊²、湯本 哲郎³、亀井 淳三³、山田 友香⁴、大野 英樹⁴ (1. 電気通信大学、2. 順天堂大学、3. 星薬科大学、4. I&H株式会社)

*Hiroki Shirakata¹, Yu Nishiyama¹, Kazutoshi Fujibayashi², Tetsuro Yumoto³, Junzo Kamei³, Yuka Yamada⁴, Hideki Oono⁴ (1. The University of Electro-Communications, 2. Juntendo University, 3. Hoshi University, 4. I&H Corporation)

関西地域 86 薬局における 抗インフルエンザ薬処方件数の予測

白方 大樹^{*1}, 西山 悠^{*1}, 藤林 和俊^{*2}, 湯本 哲郎^{*3}, 亀井 淳三^{*3}, 山田 友香^{*4},
小湊 英範^{*4}, 大野 英樹^{*4}

^{*1} 電気通信大学, ^{*2} 順天堂大学, ^{*3} 星薬科大学, ^{*4} I&H 株式会社

Prediction of Prescription Counts of Anti-influenza Drugs on 86 Pharmacies in Kansai Area.

Hiroki Shirakata^{*1}, Yu Nishiyama^{*1}, Kazutoshi Fujibayashi^{*2}, Tetsuro Yumoto^{*3}, Junzo Kamei^{*3}, Yuka Yamada^{*4}, Hidenori Kominato^{*4}, Hideki Oono^{*4}

^{*1}The University of Electro-Communications, ^{*2}Juntendo University,

^{*3} Hoshi University, ^{*4} I&H Corporation

インフルエンザは毎年約 1000 万人以上が罹患する、疫学的に重要な感染症の一つである。そのため、インフルエンザの流行予測は重要な課題である。本研究では関西地域の薬局の86店舗で処方した抗インフルエンザ薬を集計したデータを用いて抗インフルエンザ薬の処方件数を予測し各手法を比較した。予測に用いたモデルは季節自己回帰和分移動平均モデル、非線形 Support Vector Regression, ガウス過程回帰, ポアソン分布を尤度関数としたガウス過程回帰, Long Short Term Memory を用いた。比較するにあたり, Root Mean Squared Error と決定係数で評価した。その結果, LSTM の予測精度が最も高い結果になった。

キーワード インフルエンザ, 抗インフルエンザ薬, 時系列解析, ガウス過程回帰, LSTM

1. はじめに

インフルエンザは毎年約 1000 万人以上が罹患し[1], その流行状況は医療機関受診者数を基に判断される。日本の国立感染症研究所(NIID)では各都道府県ごとと保健所に報告されたインフルエンザ患者数を公開している。公開データを用いたインフルエンザ流行予測研究は様々行われてきた。機械学習モデルを比較した研究[2,3], 時空間モデルや人流を考慮した研究[4,5], Twitterを用いた研究[6]等がある。

抗インフルエンザ薬処方状況はインフルエンザ流行を反映しているとされている[7,8]。本研究では、インフルエンザ流行の代替指標として、保健所より密に存在する関西地域 86 薬局で処方された抗インフルエンザ薬処方件数のデータを用い、抗インフルエンザ薬の処方件数の時系列予測を行う。本研究では、5 つの時系列モデルを比較した：季節自己回帰和分移動平均モデル(SARIMA), 非線形 Support Vector Regression(SVR), ガウス過程回帰(GPR), ポアソン分布を尤度としたガウス過程回帰(Po-GPR), Long Short Term Memory (LSTM)。

2. 方法

1) 抗インフルエンザ薬処方データ

2013 年 1 月 1 日から 2019 年 3 月 31 日にかけて阪神調剤薬局(I&H 株式会社)の関西圏(大阪・京都・奈良・兵庫)86 店舗で、薬局単位の抗インフルエンザ薬剤(タミフル, リレンザ, ゾフルーザ)の一日処方件数の総和を「抗インフルエンザ薬処方状況」として収集した。薬局の休業日や曜日効果を除くため、週で合計した件数を用いる。機械学習モデルの訓練期間を 2013 年 1 月 1 日から 2018 年 3 月 31 日とし、テスト期間を 2018 年 4 月 1 日から 2019 年 3 月 31 日とした。入力変数は1ヶ月間(4変数), 2ヶ月間(8変数), 半年間(24変数), 一年間(52変数)を用い、いずれも一週間先の抗インフルエンザ薬の処方件数を予測した。

比較手法

以下の5つの予測モデルを比較した。表 1 は各手法の実験設定を表す。SARIMA 以外は 5-fold CV で超パラメータを決定した。

SARIMA: 季節による周期性を考慮した自己回帰和分移動平均(ARIMA)モデル。

SVR: Support Vector Machine の回帰版。

GPR: 時系列モデルにガウス過程を仮定し、予測

表 2 86 薬局の平均 RMSE と平均決定係数の結果

	1 month		2 months		6 months		1 year	
	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²
SARIMA	6.065	-15.48	6.066	-14.33	6.008	-22.61	6.083	-26.79
SVR	3.222	0.9215	3.800	0.8698	4.208	0.8459	4.545	0.8554
GPR	4.330	0.9049	4.565	0.8940	4.331	0.8933	4.373	0.8857
LSTM (layer = 150)	3.130	0.3518	3.076	0.2719	3.076	0.0765	3.007	0.1520
LSTM (layer = 300)	3.036	0.9398	2.973	0.9586	2.960	0.9565	2.936	0.9645
Po-GPR	4.380	0.9060	4.589	0.8963	4.337	0.8960	4.412	0.8845

分布をガウス分布で推定する。

LSTM:再帰型 Neural Network (RNN)の拡張であり、長い時系列データに対して学習可能な深層学習である。

Po-GPR: ガウス過程回帰の尤度関数をポアソン分布と仮定したモデルで、関数の事後分布をガウス分布に近似して学習する。

果の良かった LSTM はレイヤー数が多いほど、柔軟に関数を表現できるため最も精度が高かった。また、予測に LSTM を用いることによって特徴量生成をしなくてよいので、有用である。

参 考 文 献

[1] 厚生労働省: インフルエンザに関する報道発表資料, 2020.

[2] J. Zhang, K. Nawata : A comparative study on predicting influenza outbreaks, BioScience Trends. 2017, 11(5),533-541.

[3] L. Basile, M. Oviedo de la Fuente, N. Torner, A. Martínez, M. Jané: Real-time predictive seasonal influenza model in Catalonia, Spain.2018, PLOS ONE, 13(3), 1–15.

[4] R. Senanayake, S. O’Callaghan, and F. Ramos: Predicting Spatio-Temporal Propagation of Seasonal Influenza Using Variational Gaussian Process Regression.2016, AAAI, 3901-3907.

[5] 村山 太一, 清水 伸幸, 藤田 澄男, 若宮 翔子, 荒牧 英治: 位置関係を考慮した地域ごとのインフルエンザ流行予測,2019.

[6] E. Aramaki, S. Maskawa, M. Morita: Twitter Catches The Flu: Detecting Influenza Epidemics using Twitter, 2011, EMNLP, 1568–1576

[7] Centers for Disease Control and Prevention (CDC).Value of pharmacy-based influenza surveillance -Ontario, Canada, 2009. MMWR Morb Mortal Wkly Rep.2013 May 24;62(20):401-4.

[8] Yoshida M, Matsui T, Ohkusa Y, Kobayashi J, Ohyama T, Sugawara T, Yasui Y, Tachibana T, Okabe N.Seasonal influenza surveillance using prescription data for anti-influenza medications. Jpn J Infect Dis. 2009May;62(3):233-5.

Model	Language	Library	setting
SARIMA	R	Forecast	Auto.arima
N-SVR	R	Kernlab	ガウス RBF
GPR	Python	Gpy	ガウス RBF
Po-GPR	Python	Gpy	ガウス RBF ラプラス近似
LSTM	Python	Keras	ノード数 150,300

表 1 各手法の設定

評価指標

1. Root Mean Squared Error Loss (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_i (y_i - \hat{y}_i)^2}$$

2. 決定係数(R²)

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

y_i:観測値, $\hat{y}_i$:予測値, $\bar{y}$:平均値

3. 結果

関西地域の 86 薬局それぞれに対して抗インフルエンザ薬処方件数を各手法で予測し, 86 店舗で

RMSE とR²を平均した結果が表2である. LSTM の隠れ層のノードを300, 入力変数を52変数(一年)に設定した場合,最も良い結果が得られた. LSTM は入力変数を多くすると, 予測精度が向上した. 一方, 非線形 SVR は入力変数を少なくすると予測精度が向上した.

4. まとめと考察

抗インフルエンザ処方件数の予測を線形モデルと非線形モデル, Deep learning で比較した. 最も結