A Self-adaptive 30-day Diabetic Readmission Prediction Model based on Incremental Learning

Peng Zhao, MS[1]; Illhoi Yoo, PhD[1,2]

[1]Informatics Institute, University of Missouri [2]Department of Health Management and Informatics, School of Medicine, University of Missouri

Abstract

In the United States, hospital readmissions within 30 days of discharge account for $15 billion in Medicare expenditures annually [1]. The prediction of hospital readmissions has been a prior for researchers to identify patients at risk of early readmission. Most of the reported hospital readmission prediction models have been built with historical data and thus can outdate over time. In this work, a self-adaptive 30-day diabetic hospital readmission prediction model has been developed. A diabetic inpatient encounter data stream was used to train the self-adaptive models based on incremental learning algorithms. The result indicated that the model can automatically adapt to the newly arrived data. The best model achieved an average area under that ROC curve (AUC) of 0.655 ± 0.078.

Introduction

Hospital readmission: a patient is hospitalized again after being discharged from a hospital within a short period (normally 30 days).

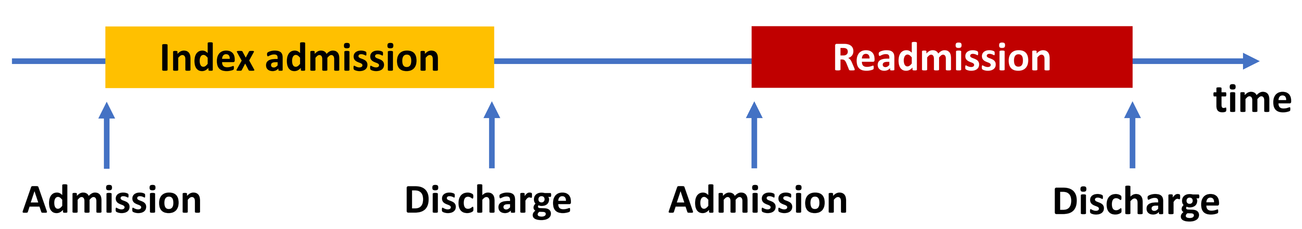

Figure 1. The temporal relationship between an index admission and a readmission.

Traditional supervised learning: assumes the training data is complete and generates static prediction models, which can become outdated because the distribution of readmissions evolves over time. To keep up-to-date, these models need to be re-trained periodically to incorporate new data. It can be expensive and hard to determine and discard outdated knowledge manually.

Data stream learning: continuously updates a prediction model on-the-fly with newly arrived data instead of rebuilding it from scratch.

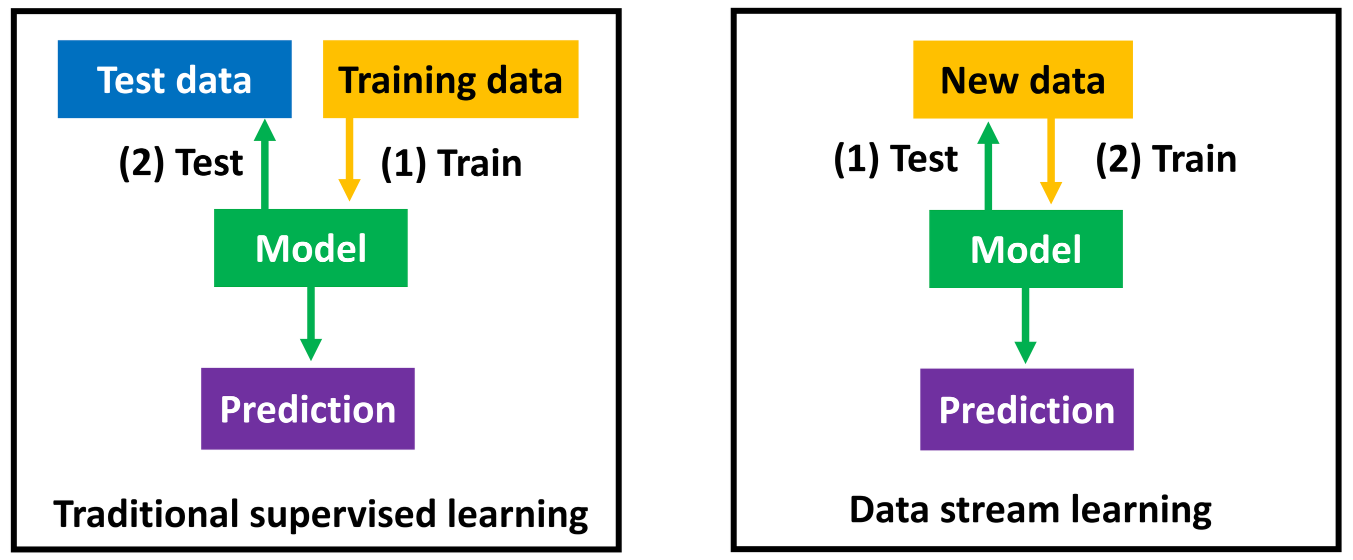

Figure 2. A comparison of traditional supervised learning and data stream learning processes.

30-day diabetic readmissions were studied because the 30-day readmission rate of diabetic patients has been higher than the overall 30-day readmission rate (14.4-22.7% vs. 8.5-13.5%) [2,3] but has drawn relatively less attention.

Methods

Data: 101,766 de-identified inpatient diabetic encounters from 130 US hospitals during 1999-2008 with 11.2% being readmitted within 30 days after discharge [4]. The data contains 50 clinical and administrative variables. The dimension has reduced to 22,132 × 48 after handling missing values and balancing the 2 classes (readmitted within 30 days or not).

Online Naïve Bayes [5]:

$$P_k(\theta) = P(\theta|x_k) = \frac{P(x_k|\theta)P_{k-1}(\theta)}{P(x_k)}$$

Hoeffding Tree [6]: choose a as the split attribute if: $G(a) - G(b) > \sqrt{\frac{R^2 ln\left(\frac{1}{\delta}\right)}{2n}}$

a, b	Attributes with the best and second best G
G	Information gain of choosing an attribute
R	Range of observed independent instances on a leaf
n	Number of observed independent instances on a leaf
$1-\delta$	Confidence

a) It uses a small number of instances to choose the split attribute.
b) The output is nearly identical to that of a non-incremental decision tree

Hoeffding Adaptive Tree [7]:

a) It is based on Hoeffding Tree and considers concept drift.
b) It monitors the performance of tree branches and replaces them with more accurate new branches if their accuracy decreases.

Results

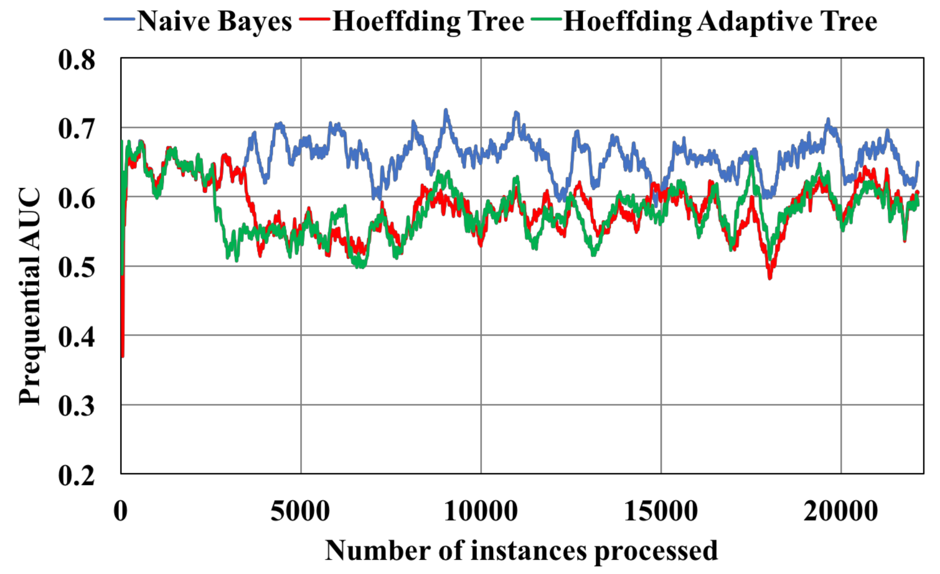

Figure 3. The AUC curves of the three models. The monitoring window size is 500.

Figure 3 shows the AUC curves of three candidate models. As the monitoring window (n = 500) moves along the data stream, the classifications within the window will be used to calculate the AUC. It can be seen that the online Naïve Bayes model outperforms the other two tree models with the AUC fluctuating between 0.6 to 0.75 and a mean of 0.655 ± 0.078.

Results

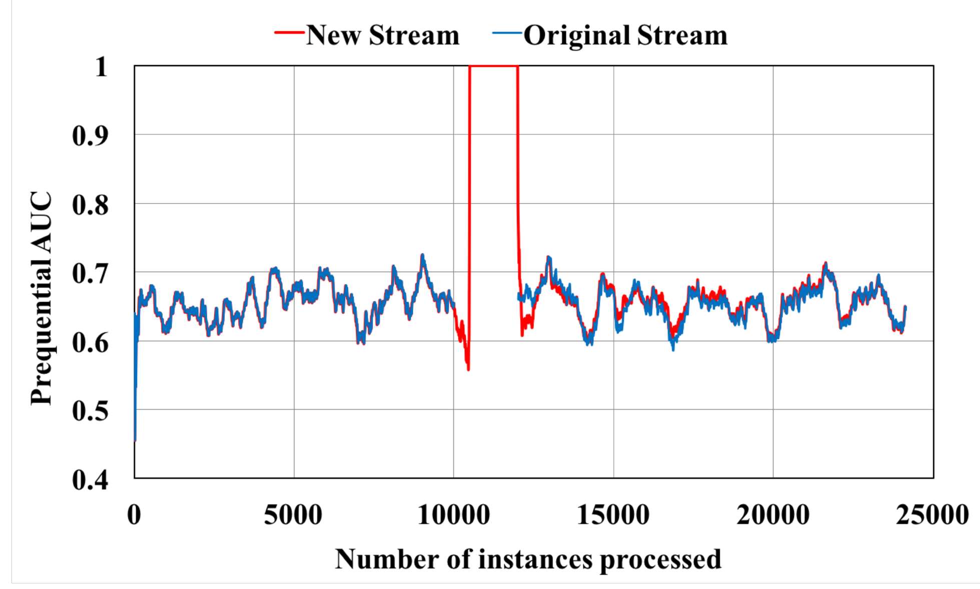

Figure 4. The AUC curves of the online Naïve Bayes models for the modified data stream and the original data stream. The monitoring window size is 500.

To test the online Naïve Bayes model's ability to adapt to new data, the original data stream was modified by inserting 2,000 new data instances after the 10,000th instance. These 2,000 instances are of the same class and have never been seen by the model. The result is shown in Figure 4. As the window moves along the data stream, it will enter a region where all the instances are of the same class and the AUC will be equal to 1. After the window has moved out the single class region, the AUC becomes less than 1 and the two curves have very high degree of agreement. The minor disagreements between 12,000 and 19,000 indicate the model retains the memory of the inserted 2,000 instances in the modified data stream. After 19,000, the model gradually adapts to the new trend and forgets the outdated knowledge of the inserted 2,000 instances.

Conclusion

In this work, we have developed a self-adaptive 30-day diabetic hospital readmission prediction model based on incremental learning. Compared to the static models that can get outdated over time, the presented incremental model can constantly update itself when new data becomes available. The model has been proven to be able to retain the memory of previous data for a period of time and gradually forget the outdated knowledge and adapt to the latest trend.

Reference

1. "Report to the Congress: promoting greater efficiency in Medicare", Washington, DC : Medicare Payment Advisory Commission (MEDPAC), 2007.
2. "Economic costs of diabetes in the U.S. in 2012", *Diabetes Care*, vol. 36, no. 4, pp. 1033–1046, Apr. 2013.
3. J. M. Robbins et al, "Diagnosing diabetes and preventing rehospitalizations: the urban diabetes study," *Medical care*, vol. 44, no. 3, pp. 292–6, Mar. 2006.
4. B. Strack et al, "Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records," *BioMed Research International*, vol. 2014, pp. 1–11, 2014.
5. T. Bayes et al, "An essay towards solving a problem in the doctrine of chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S.," *Philosophical Transactions of the Royal Society of London*, vol. 53, pp. 370–418, Jan. 1763.
6. G. Hulten et al. "Mining time-changing data streams," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01*, 2001, pp. 97–106.
7. A. Bifet et al. "Adaptive learning from evolving data streams," in *IDA 2009: Advances in Intelligent Data Analysis VIII*, 2009, pp. 249–260.

Contact
Peng Zhao
Email: peng.zhao@mail.missouri.edu
Tel: +1.801.831.8843

Informatics Institute
University of Missouri

Health Management and Informatics
School of Medicine