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

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

[1]Informatics Institute

[2]Department of Health Management and Informatics, School of Medicine

University of Missouri

Columbia, MO, 65211, USA

University of Missouri

HOSPITAL READMISSION

Definition:

- A patient is hospitalized again after being discharged from an index (initial) admission within a specified time window.

HOSPITAL READMISSION

High frequency and high associated cost:

- In the United States, the 30-day Medicare unplanned readmission rate was **15.2%** during July 2015 to June 2016 [1].
- In England, the 28-day emergency readmission rate was **11.5%** during the fiscal year 2011 to 2012 [2].
- In 2011, hospital readmissions contributed **$41.3 billion** in **total** hospital costs in the United States [3].
- Several countries have implemented programs to financially penalize readmissions, such as England, Germany, and the United States [4].

HOSPITAL READMISSION

Predictive modeling

- Numerous predictive models have been built to:
  - Predict a patient's readmission risk.
  - Predict a hospital's readmission rate.
- Problems:
  - In two systematic reviews of **99** readmission predictive models reported between 1985 and 2015, **ALL** the models are **STATIC** [5,6].
  - The static models can get outdated because the distribution of the patient population can evolve over time and it is hard to determine the "shelf life" of the model.
  - To keep up-to-date, the models need to be re-built from scratch constantly, which can be expensive in practice.

OBJECTIVE

- Build a self-adaptive 30-day readmission prediction model, which can update itself when new data become available.
  - Incremental learning: adapt to new data and extend the existing model's knowledge without retraining.

DATA & PREPROCESSING

- A de-identified inpatient encounter dataset of diabetic patients from 130 US hospitals during 1999-2008 [7].
    - 101,766 inpatient encounters
    - 11.2% encounters with 30-day readmissions
    - 50 attributes
- The dataset was balanced by down sampling the major class (encounters without 30-day readmissions).
- 22 attributes were selected.

TOOLS & ALGORITHMS

Massive Online Analysis

- Naïve Bayes [8]
- Hoeffding Tree (VFDT) [9]
- Hoeffding Adaptive Tree [10]

PREQUENTIAL EVALUATION

- When new data instances become available, they are first used to test the model and then used to improve the model [11].
- Prequential AUC: only evaluates the classifications of the most recent instances within a sliding window of size K [12].

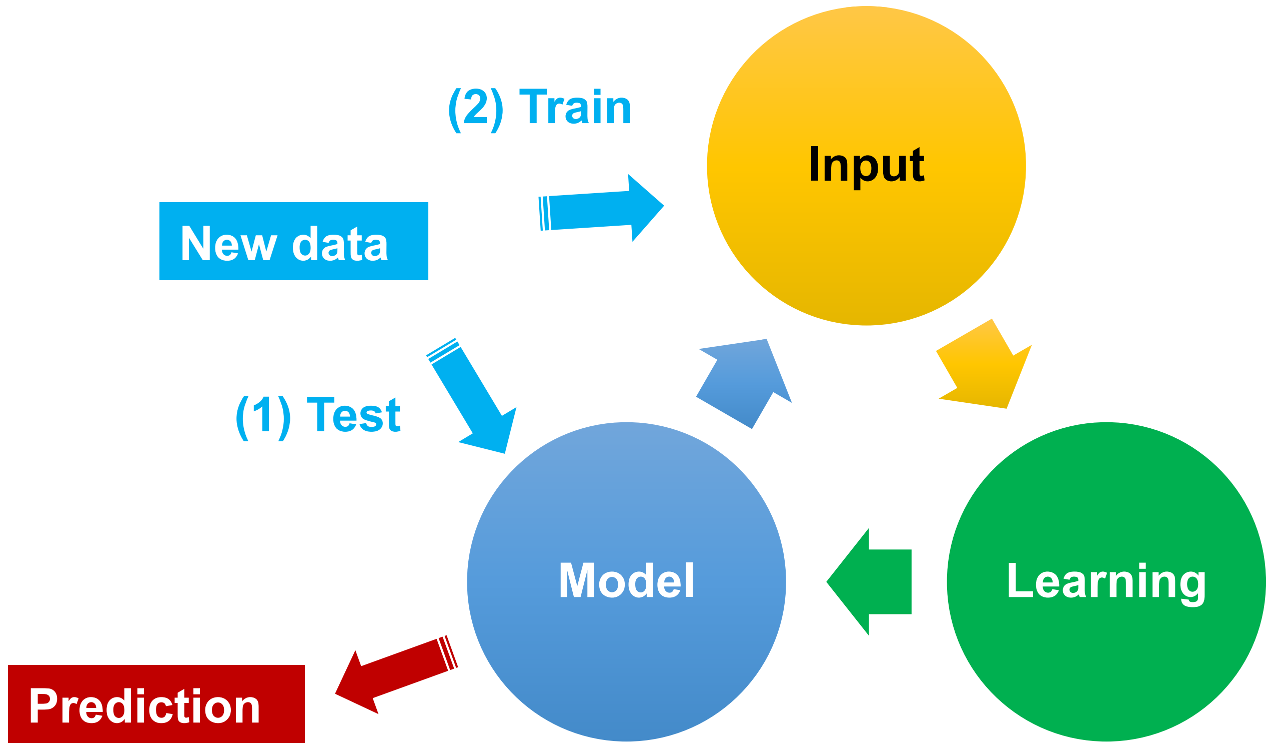
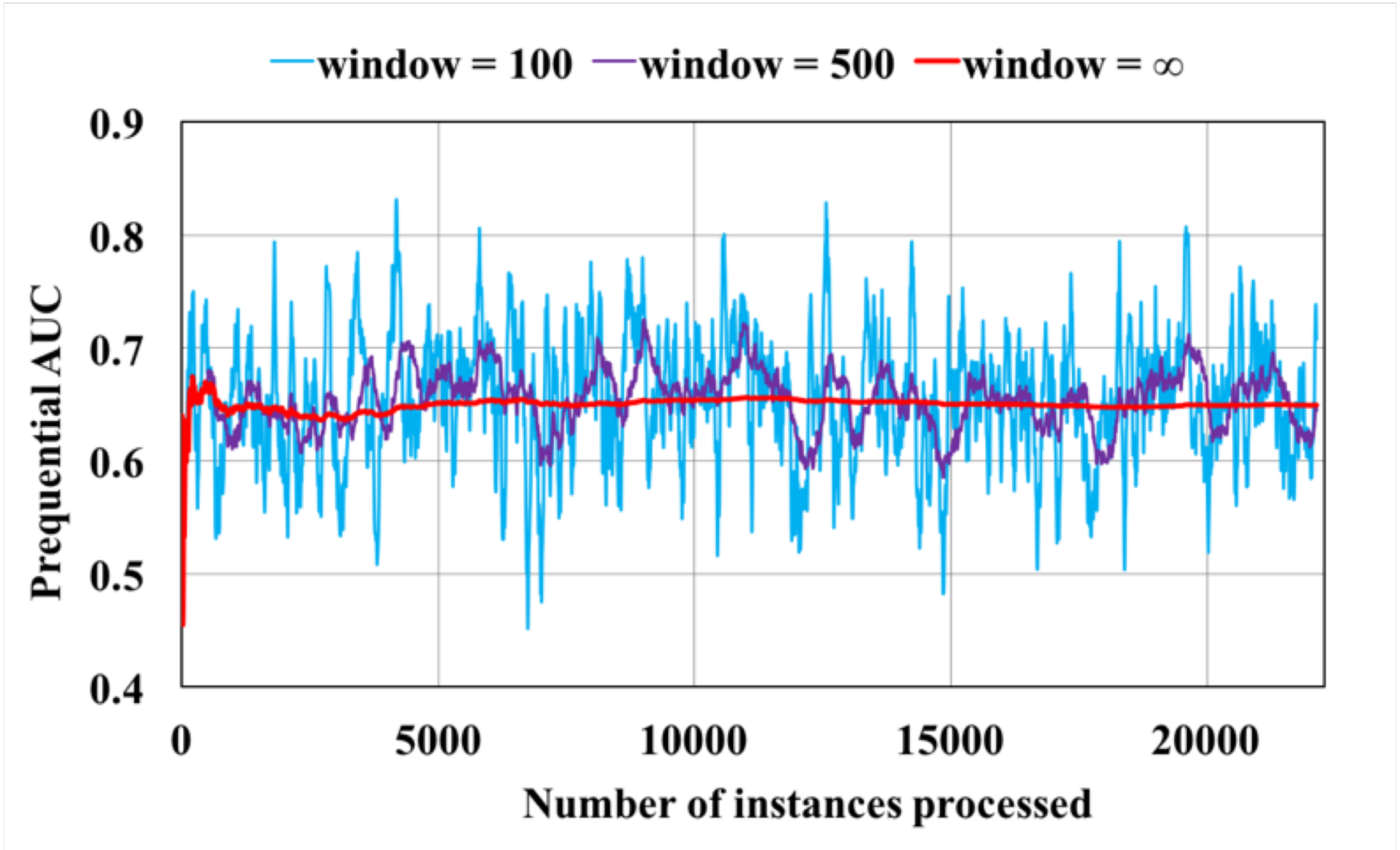


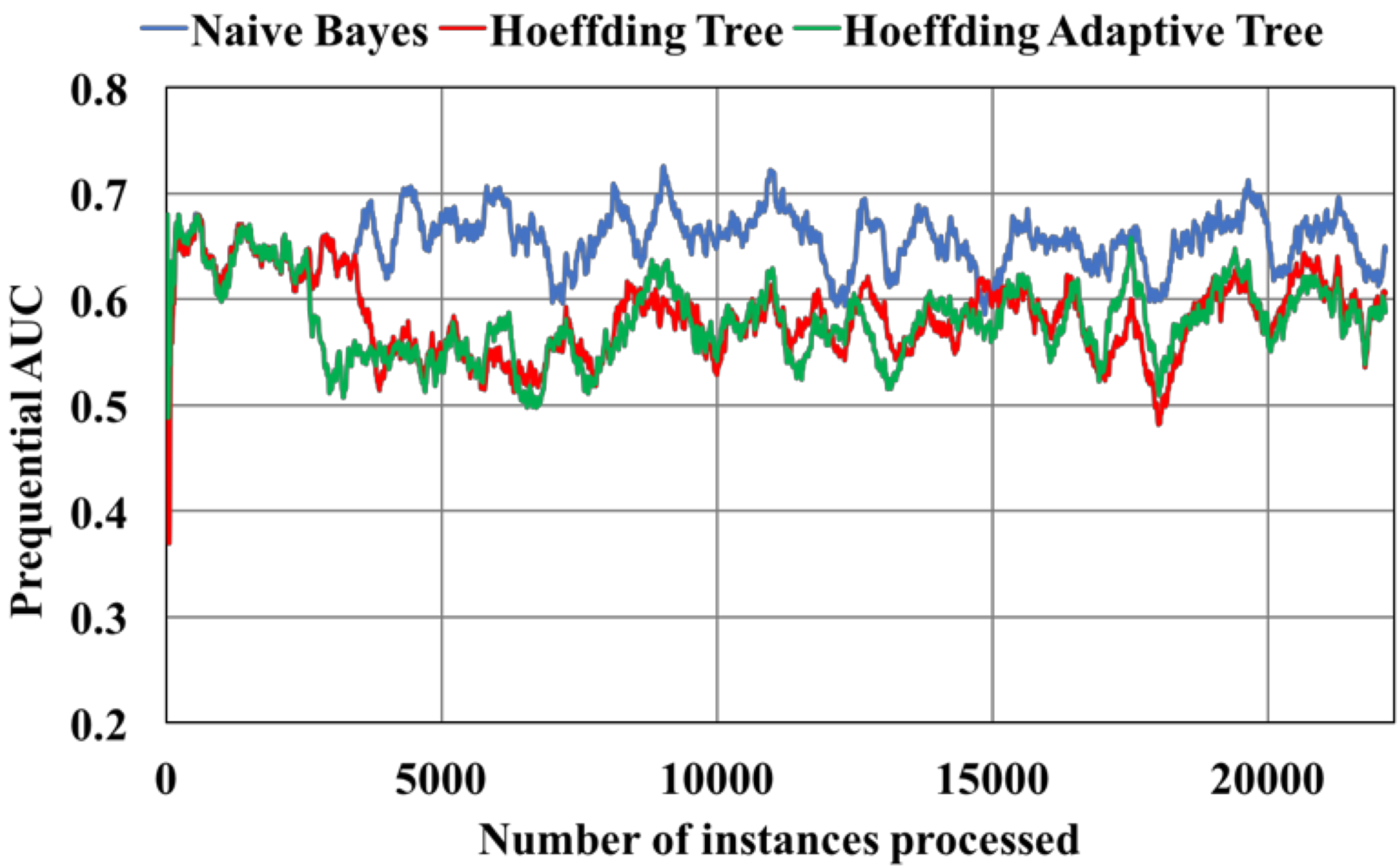
Image adapted from: https://moa.cms.waikato.ac.nz/tutorial-1-introduction-to-moa/

RESULTS



Graph 1. Effects of the sliding-window size K on the prequential AUC curves of the Naïve Bayes model.

RESULTS



Graph 2. The prequential AUC curves of the three algorithms. K = 500

RESULTS

- Benchmark: compare the performance of the self-adaptive model with a static model.

Table 1. The AUC of the incremental and static models trained with the same data.

Models	AUC
Naïve Bayes (incremental)	0.655 ± 0.078 (average)
Naïve Bayes (static)	0.660
Naïve Bayes (static) [13]	0.657

CONCLUSION

- A self-adaptive 30-day diabetic hospital readmission prediction model has been developed based on incremental learning.
- The global performance of the incremental model is nearly identical to that of the static models built with the same dataset.

ACKNOWLEDGEMENT

- PhD advisor:
  - Illhoi Yoo, PhD
- PhD committee:
  - Abu Mosa, PhD
  - Mihail Popescu, PhD
  - Suzanne Boren, PhD

Questions?

peng.zhao@mail.missouri.edu

REFERENCES

1. Hospital Quality Initiative - Outcome Measures 2017 Chartbook https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/OutcomeMeasures.html
2. HSCIC. Hospital Episode Statistics, Emergency readmissions to hospital within 28 days of discharge - Financial year 2011/12. 2013.
3. Hines AL, Barrett ML, Jiang J, Steiner CA. Conditions with the Largest Number of Adult Hospital Readmissions by Payer, 2011. Healthcare cost and Utilization Project 2014;363(7):1–10
4. Kristensen SR, Bech M, Quentin W. A roadmap for comparing readmission policies with application to Denmark, England, Germany and the United States. Health Policy Elsevier Ireland Ltd; 2015;119(3):264–273. PMID: 25547401
5. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, et al. Risk prediction models for hospital readmission: a systematic review. JAMA 2011 Oct 19;306(15):1688–98.
6. Zhou H, Della PR, Roberts P, Goh L, Dhaliwal SS. Utility of models to predict 28-day or 30-day unplanned hospital readmissions: an updated systematic review. BMJ Open 2016 Jun 27;6(6):e011060.
7. 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.
8. P. Domingos and M. Pazzani, On the Optimality of the Simple Bayesian Classifier under Zero-One Loss, Machine Learning, vol. 29, pp. 103–130, 1997.
9. G. Hulten, L. Spencer, and P. Domingos, 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.
10. Bifet and R. Gavaldà, Adaptive Learning from Evolving Data Streams, in IDA 2009: Advances in Intelligent Data Analysis VIII, 2009, pp. 249–260.
11. http://moa.cms.waikato.ac.nz/
12. Brzezinski et al. Prequential AUC for classifier evaluation and drift detection in evolving data streams. Proceedings of the 3rd International Workshop on New Frontiers in Mining Complex Patterns, Nancy, France. 2015, 87-101.
13. T. M. Ahmed, Using Data Mining To Develop Model for Classifying Diabetic Patient Control Level Based on Historical Medical Records, Journal of Theoretical and Applied Inforamtion Technology, vol. 87, no. 2, pp. 316–323, 2016.

ALGORITHMS

Naïve Bayes

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

- Naturally incremental.
- Constantly updates counts according to incoming training instances.

ALGORITHMS

Hoeffding Tree (VFDT)

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

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

[1] Hulten et al. Mining time-changing data streams. ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, 2001 97-106
[2] Zang et al. Comparative study between incremental and ensemble learning on data streams: Case study. Journal of Big Data, 2014, 1:5

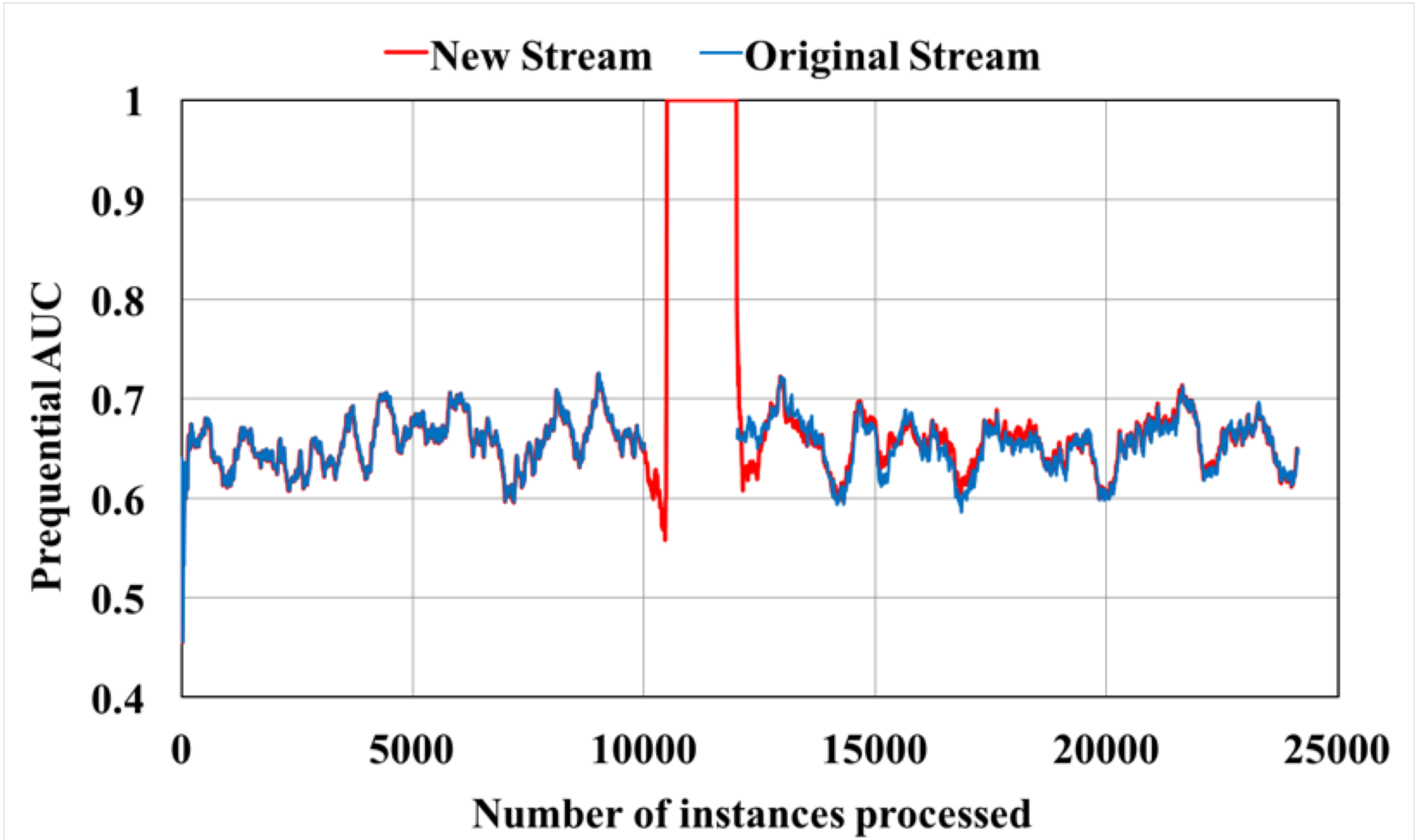
ALGORITHMS

Hoeffding Adaptive Tree

- Based on Hoeffding Tree.
- It monitors the performance of branches and replaces them with more accurate new branches if their accuracy decreases.

[1] Bifet et al. Adaptive learning from evolving data streams. Ricard Gavaldà. IDA 2009: Advances in Intelligent Data Analysis VIII pp 249-260

RESULTS



Graph 3. The prequential AUC curves of the Naïve Bayes models for the modified data stream and the original data stream. K = 500