The Thirty-Ninth AAAI Conference on Artificial Intelligence (AAAI-25)

Robust and Adaptive AI Models for Medication Usage Forecasting Using ICD-9/10 Code (Student Abstract)

Jonathan Li
McLean High School, 1633 Davidson Rd, McLean, VA 22101 USA
jony.li2009@gmail.com

Abstract

Accurate forecasting of medication usage and ICD-9/10 code streams is critical for optimizing medical logistics, especially during periods of high demand, such as pandemics, disease outbreaks, wartime, or natural disasters. In this study, we develop a novel and robust forecasting framework using unsupervised learning techniques and Natural Language Processing (NLP) methods to build vector representations of daily ICD-9/10 codes and medication daily usage from Electronic Health Record (EHR) data. Multiple forecasting models, including Linear Drift Model, Vector Autoregression (VAR), Temporal Fusion Transformer (TFT), and Autoregressive Long Short-Term Memory (AR-LSTM) are trained, tested and evaluated. Finally multiple TFT and AR-LSTM models with different lookback horizon are trained and ensembled together to achieve better forecasting accuracy in near further (10 days). The AI framework is validated using MIMIC-IV ER and MIMIC-III datasets, resulting in the average forecasting error 5.2% at 5-th day and 18.1% at the 10-th day. The results demonstrate the ensemble model's superior performance on near-future medication usage forecasting and ICD code progression, offering valuable insights for healthcare logistics and decision-making. The framework also provides the mechanism to detect the model drift and finetune the model if necessary, which offers a robust tool for managing healthcare logistics under extreme and fluctuating conditions.

Introduction

The COVID-19 pandemic has highlighted the critical importance of efficient management and forecasting of medical resources, including medications and other essential medical supplies. During the pandemic, many healthcare systems worldwide faced severe shortages of medications, and other vital medical resources, resulting in tragic outcomes, including the loss of many lives (Callaway Kim et al. 2024). The inability to accurately predict medication and medical supply usage, and anticipate supply chain disruptions exposed vulnerabilities in healthcare logistics and underscored the need for robust and reliable forecasting models.

To address such vulnerabilities, a combination of unsupervised learning methods and Incremental Principal Component Analysis (PCA) (Ross et al. 2008), along with Natural Language Processing (NLP) techniques, are employed to create meaningful vector representations of the ICD-9/10 code sequences and medication usage patterns extracted from EHR data, which are further compressed into lower-dimensional vectors for various machine learning models, including latest deep learning ones. Several machine learning models are trained and evaluated from the data in this study. To achieve better forecasting performance, an ensemble method is developed to combines the outputs from multiple models by leveraging the complementary capabilities of different architectures. To handle the stochastic and non-static characteristics of the time series, Bayesian change-point detection algorithm (Adams and MacKay 2007) is applied to detect the model drift. The system will retrain the model weighting more on latest historical data if drift occurs, which also provides the capability to adapt the emergence situations, such as pandemic, outbreak or the war zones.

Methodology

The framework mainly forecasts the medication usage from patient stream, encoded by ICD 9/10 codes. The input is the ICD 9/10 occurrence daily stream in the hospitals or medical facilities; and the output is the daily usage for each medication in near future (10 days). It has four main modules: ICD Encoder, Medication Usage Encoder, Forecasting Models, and Stratified Error Calculation and Model Drift Detection.

Figure 1 shows the detail how to build the ICD encoding base and vector daily point for ICD code occurrence. All ICD 9 codes are converted into ICD 10 code, and grouped by first 3 characters, which provide the fundamental grouping for diseases, conditions, or external causes of injury, and subsequent characters. The incremental principal component analysis (Ross et al. 2008) is applied to reduce the dimensionality and keep 97% variation of original vector space. Small regularization Gaussian noise is added to the data during projection to improve the numerical stability and avoid rank deficiency.

Figure 2 shows the detail how to build the medication vector basis from FDA full medication list, where active ingredients are used to model same medication with different names. The dosage is factored into usage with normalized unit. TF-IDT (Term Frequency-Inverse Document Frequency) is applied to build the vector representation for the ingredients, and incremental PCA and K-means are used to

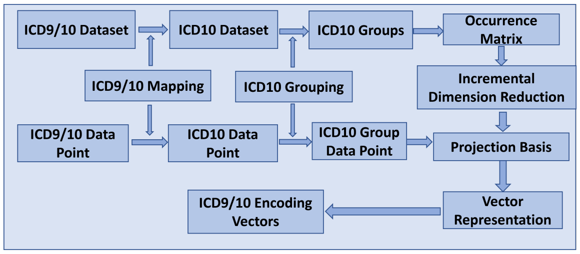

Figure 1: ICD Code Occurrence Encoder

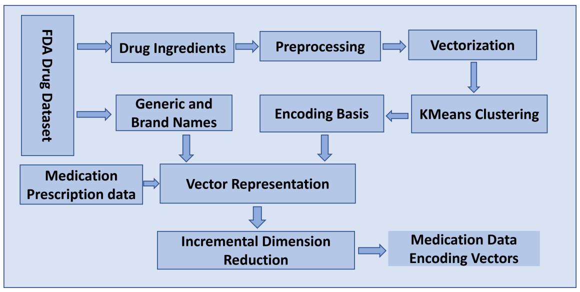

Figure 2: Medication Usage Encoder

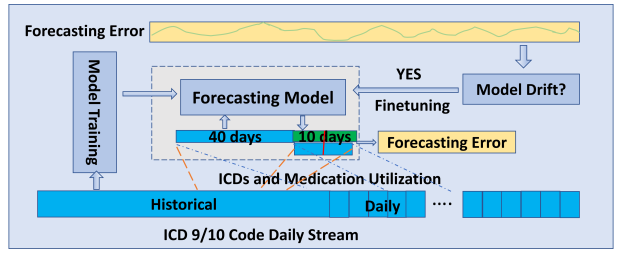

Figure 3: Drift Detection and Model Updating

Model / Med	VAR	TFT	AR-LSTM	Ensemble
Pseudoephedrine	23/30	30/30	28/30	30/30
Cepacol Lozenges	64/32	31/32	29/32	31/32
Prednisone	18/25	20/25	22/25	23/25
Esomeprazole Magnesium	45/36	40/36	37/36	38/36
Benzoyl Peroxide	32/51	49/51	53/51	52/51

Table 1: Medication Usage Forecasting at 5-th day

reduce the computational cost for deep learning forecasting models.

Several machine learning models are trained and evaluated in this study: Drift Linear model, Vector Autoregression Model, Temporal Fusion Transformer (TFT) (Lim et al. 2020) and Autoregressive-LSTM (AR-LSTM). The latter two capture both short-term and long-term dependencies within time series, outperform VAR and Drift model a lot, especially trained with Stochastic Weight Averaging (SWA)(Izmailov et al. 2019). The models are further trained with different lookback horizons and ensembled together to achieve better forecasting accuracy and reliability.

Single error metric, such as MAE and RMSE, is not suitable to evaluate the model performance on time series predictions. In this project, stratified error analysis with dynamic time warping (DTW) distance(Giorgino 2009) is used to compute error metrics in different quantiles, providing a more detailed understanding of model performance on different parts of the data distribution, revealing potential weaknesses in the model for optimization.

To handle model drift, Bayesian Changepoint Detection (Adams and MacKay 2007) is applied to detect when forecasting error increases significantly, which triggers the model finetuning weighting more on latest data points, as shown in 3. It provides a robust way to handle non-static, stochastic and dynamic characterization of time series.

Results and Conclusion

The AI forecasting framework is trained, tested and evaluated on the MIMIC-IV ER and MIMIC-III datasets. Both TFT and AR-LSTM are trained with SWA and different lookback horizons (20, 30, 40 and 50 days), and ensembled with median method, thus the final model adapts dynamically to both short-term fluctuations and long-term trends in medication usage. Table 1 shows the forecasting results for several medications, and figure 4 shows the examples of medication usage prediction, ground-truth, and the corresponding top ICD10 group occurrence.

Overall the final ensemble model perform well on many medications; the mean error for the 5-th day forecasting is 5.2% and the mean error for the 10-th day forecasting is as low as 18.1%. However, the performance indeed varies much with different medications; the maximal error of 10-th day forecasting is more than 50% for some medications. We are examining the detail and trying to find the insight for such poor performance, and improve the robustness of the framework. Bayesian Changepoint Detection algorithm is applied in the project to detect if the model performance drifts significantly. If it happens, the framework finetune the models with high sampling weights on latest data points such than the new model will capture the new pattern for better forecasting. The framework can be easily adapted to different time scale for different requirement, and provides healthcare providers with actionable insights for inventory and distribution planning.

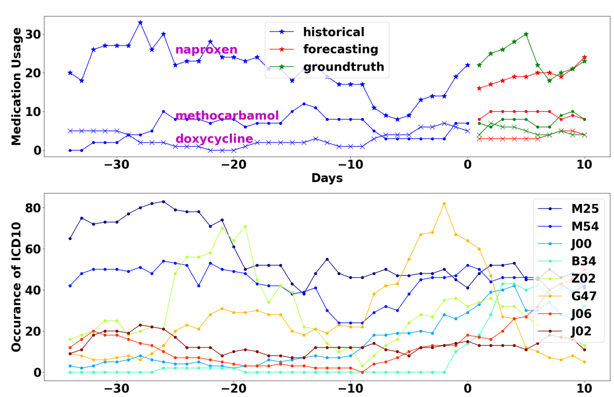

Figure 4: Medication Usage and ICD 10 Group Forecasting

Acknowledgments

In here, the author express his gratitude to Mr. Fayin Li for his help building the Python packages to extract, preprocess the data and train the model. He also provides huge insightful feedback on model improvement and optimization. Without his help, this project would not have turned out as well as it has now.

References

Adams, R. P.; and MacKay, D. J. C. 2007. Bayesian Online Changepoint Detection. arXiv:0710.3742.

Callaway Kim, K.; Rothenberger, S. D.; Tadrous, M.; Hernandez, I.; Gellad, W. F.; Devine, J. W.; Hershey, T. B.; Maillart, L. M.; and Suda, K. J. 2024. Drug Shortages Prior to and During the COVID-19 Pandemic. *JAMA Network Open*, 7(4): e244246–e244246.

Giorgino, T. 2009. Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *Journal of Statistical Software*, 31(7): 1–24.

Izmailov, P.; Podoprikhin, D.; Garipov, T.; Vetrov, D.; and Wilson, A. G. 2019. Averaging Weights Leads to Wider Optima and Better Generalization. arXiv:1803.05407.

Lim, B.; Arik, S. O.; Loeff, N.; and Pfister, T. 2020. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. arXiv:1912.09363.

Ross, D.; Lim, J.; Lin, R.; and Yang, M. 2008. Incremental learning for robust visual tracking. *International Journal of Computer Vision*, 77(1-3): 125–141.