Continually-Adaptive Representation Learning Framework for Time-Sensitive Healthcare Applications

Akash Choudhuri, Hankyu Jang, Alberto M. Segre, Philip M. Polgreen, Kishlay Jha, Bijaya Adhikari

University of Iowa

Funded by CDC MInD Healthcare Network Grant

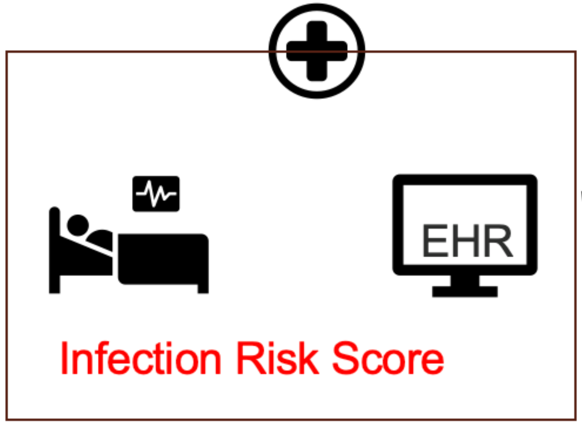
University of Iowa
comp|epi
computational epidemiology research

CDC
Centers for Disease Control and Prevention

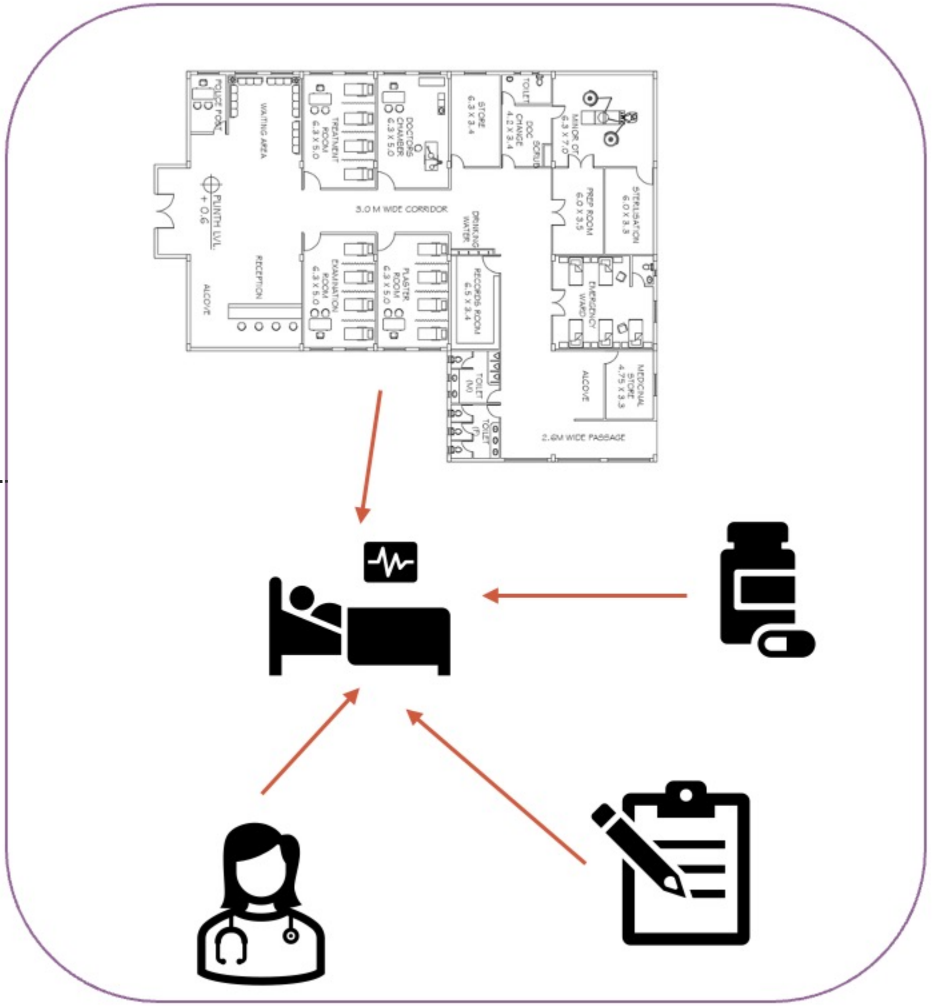
The University of Iowa

Motivation 1: Learning representations of Patients

Motivations and Principle

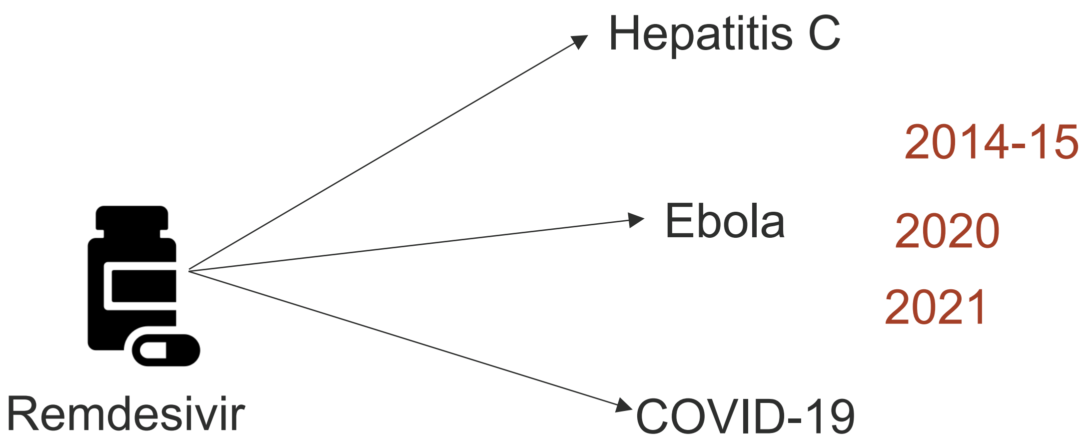


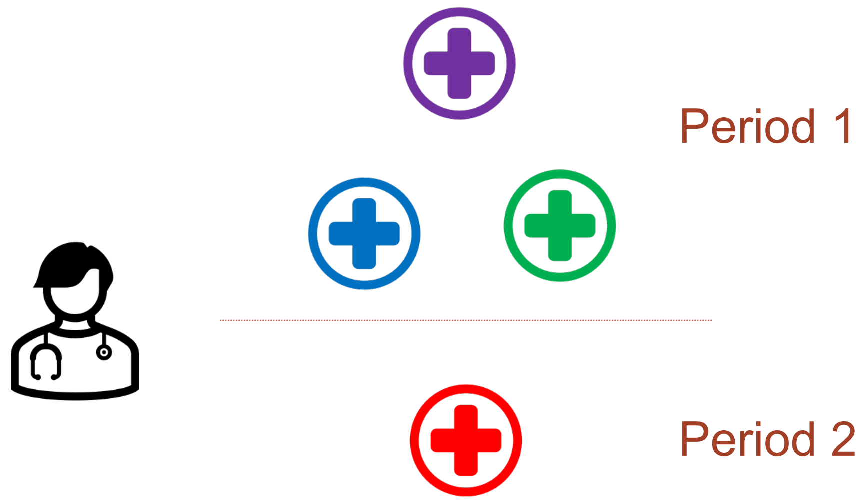
Patient Features	Outcome



Motivation 2: Incremental Incorporation of New Information

Motivations and Principle



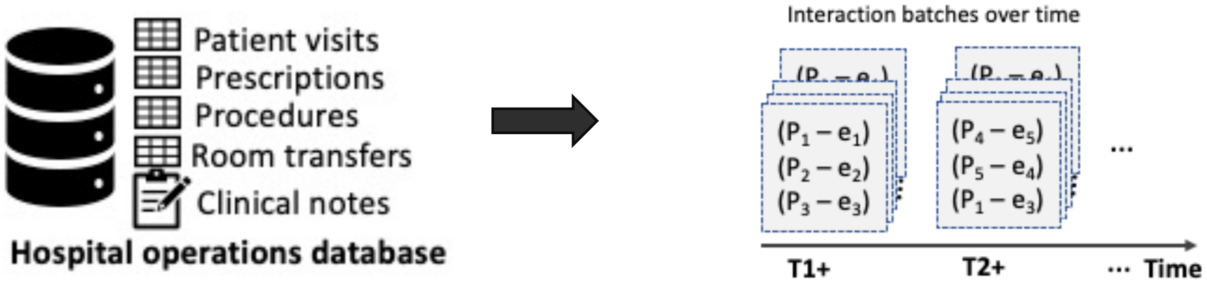


Faster Training!

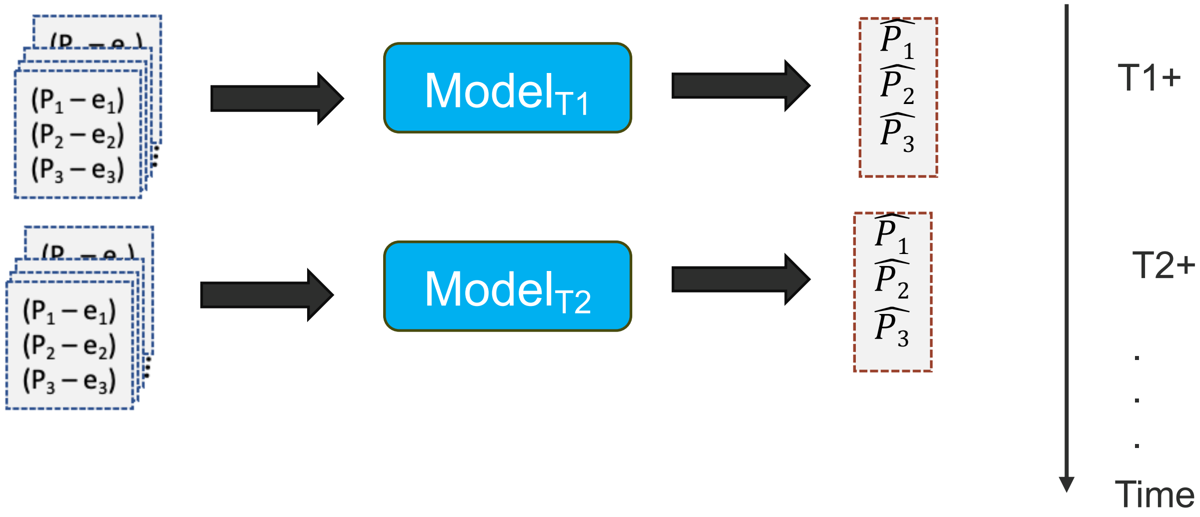
Problem Formulation

Model and Components

Given:



Learn:

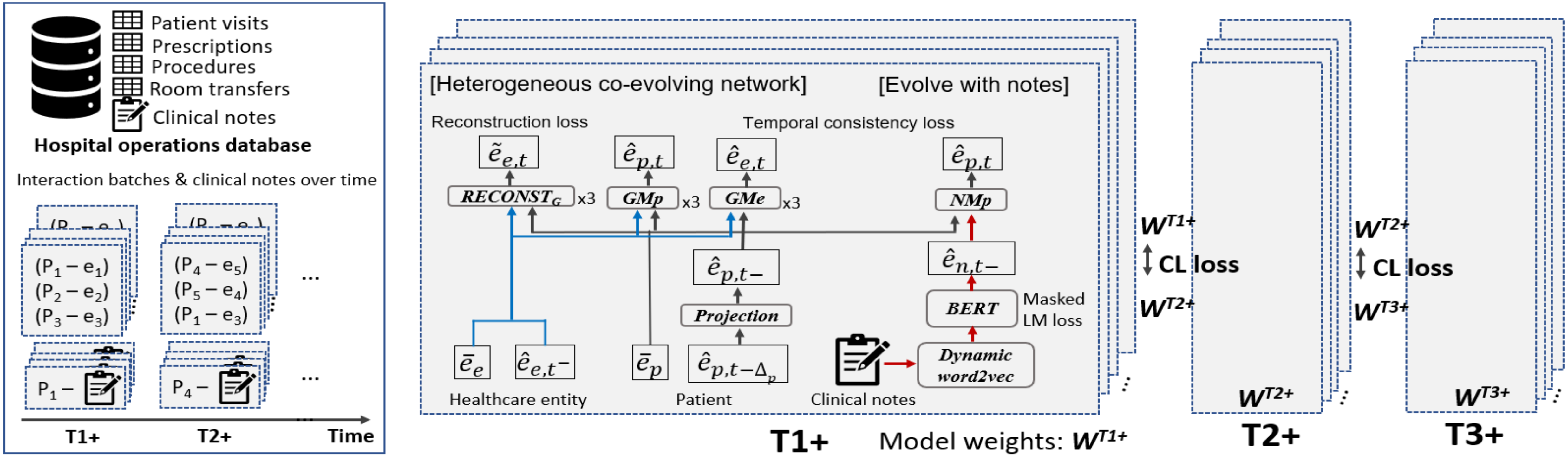


Such That:

- Dynamic patient embeddings encodes information to aid predictions
- The model parameters across periods doesn't drastically change

Model Architecture

Model and Components

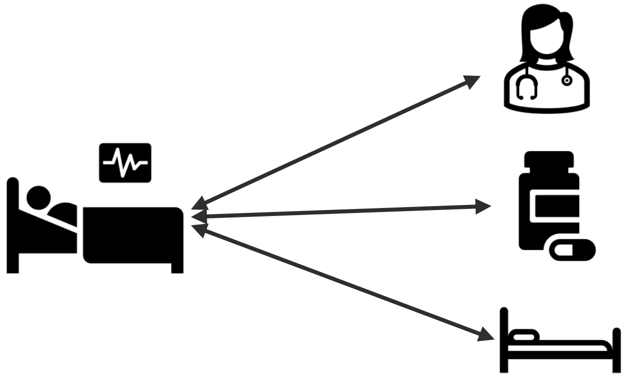


General purpose, unsupervised and continually learning embedding method for dynamic heterogenous interactions

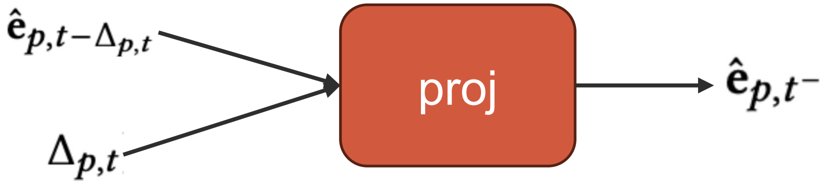
- Preserves information on the interaction via interaction type specific autoencoder
- Continually infuses knowledge across periods to prevent catastrophic forgetting

Dynamic Embedding Update

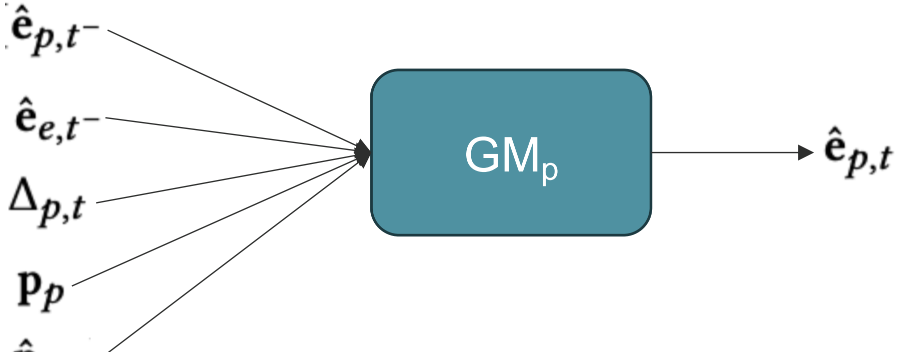
Model and Components

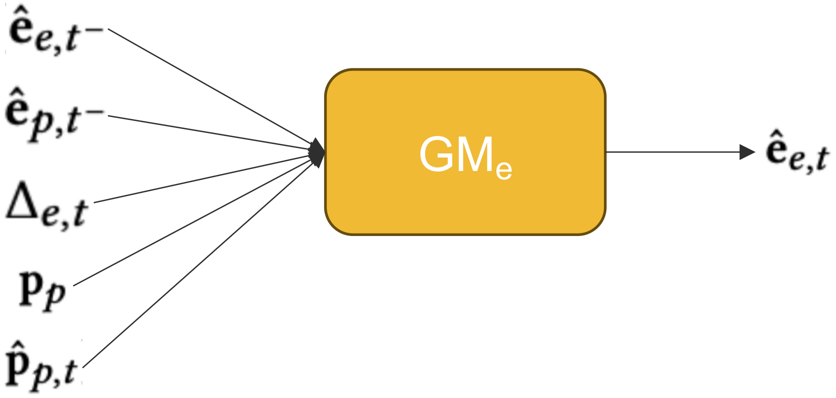


- Projection of Patient embedding (from time $t - \Delta$ to t^-) [1]:



- Update dynamic embeddings of patient and the entity at t via co-evolving neural networks:





$$\mathcal{L}_{temp} = \sum_{(p,e,t)\in S} ||\hat{e}_{p,t} - \hat{e}_{p,t^-}||_2 + ||\hat{e}_{e,t} - \hat{e}_{e,t^-}||_2$$

[1] S. Kumar, X. Zhang, and J. Leskovec, "Predicting dynamic embedding trajectory in temporal interaction networks," in ACM SIGKDD, 2019

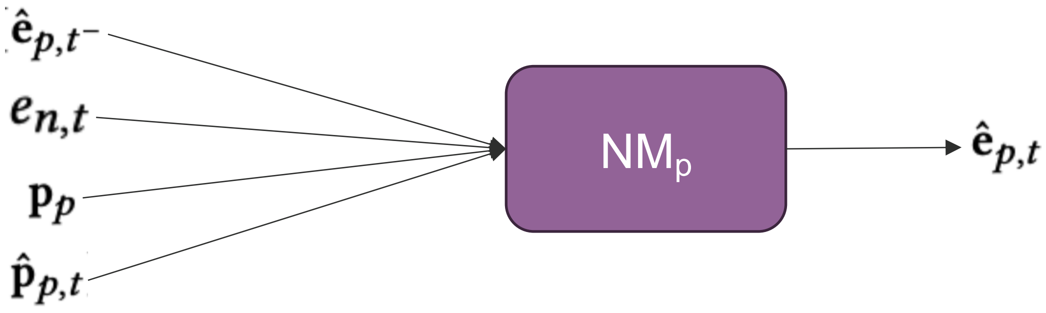
Evolution with Clinical Notes

Model and Components

- Obtaining Clinical Note Embeddings
  - For each clinical note in a period, obtain learned word embeddings using DynamicWord2Vec[1]
  - Use learned word vector embeddings to pre-train BERT[2] on Masked Language Loss:

$$\mathcal{L}_{LM} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)]$$

- Update Dynamic Patient embeddings:

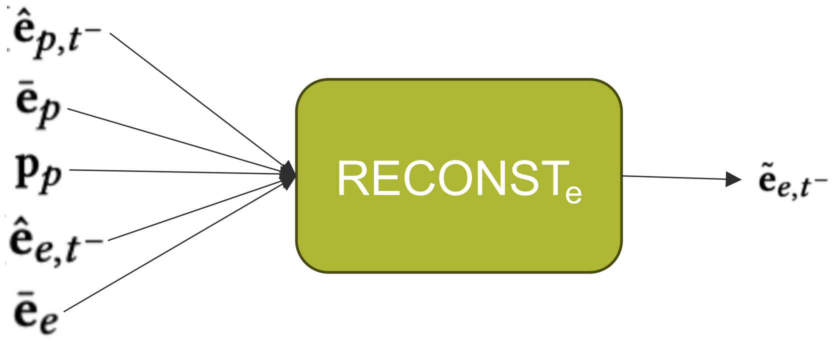


[1] Z. Yao, Y. Sun, W. Ding, N. Rao, H. Xiong, “Dynamic Word Embeddings for Evolving Semantic Discovery” in ACM WSDM, 2018
[2] J. Devlin, M. Chang, K. Lee, K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding” in arXiv, 2019

Reconstruction and Continual Knowledge Infusion

Model and Components

- We reconstruct the original entity dynamic and static embeddings via a reconstruction autoencoder:



- For each period, we prevent ‘catastrophic forgetting’ across periods by:
  - Initializing model parameters for a new period with the learned model parameters from the previous period
  - Minimizing the Continual Learning loss:

$$\text{Model}_{T1} \approx \text{Model}_{T2} \approx \text{Model}_{T3} \quad \ldots$$

$$\mathcal{L}_{CL} = \lambda||\theta_i - \theta_{i-1}||_2$$

Data

- Hospital Operations Data was obtained from University of Iowa Hospitals and Clinics (UIHC) data on:
  - **Electronic Health Records**
  - **Admission- Discharge-Transfer (ADT) logs**
- Hospital Operations was divided into 3 periods:

Period	Start Date	End Date	No. of D,M,R Interactions	No. of N Interactions
Period 1	5/4/2008	6/25/2008	245,043	149,685
Period 2	6/13/2008	8/7/2008	252,089	152,037
Period 3	7/10/2008	8/31/2008	257,994	163,158

- **Assumption:** No new entities are added across the periods

THE UNIVERSITY OF IOWA

CDI Incidence Prediction

Results

- Clostridioides difficile infection (CDI) is one of a common HAI, increases mortality risk of patients with weakened immune system
- Binary Classification Problem:
  - Instance: Patient at time t and features at that time
  - Label: Binary indicator of getting infection in next 3 days[1]
- Evaluation Metric: ROC-AUC Score
- 3- fold cross validation with 30 repetitions

Period	Method	LR	SVM	RF
Period 1	DOMAIN	0.49 ± 0.20	0.52 ± 0.07	0.34 ± 0.07
	JODIE	0.44 ± 0.12	0.36 ± 0.09	0.52 ± 0.03
	DECENT	0.62 ± 0.07	0.57 ± 0.01	0.61 ± 0.06
	Ours	$\mathbf{0.65 \pm 0.05}$	$\mathbf{0.60 \pm 0.04}$	$\mathbf{0.73 \pm 0.07}$
Period 2	DOMAIN	0.60 ± 0.11	0.54 ± 0.13	0.76 ± 0.19
	JODIE	0.50 ± 0.05	0.47 ± 0.06	0.52 ± 0.18
	DECENT	0.71 ± 0.02	0.59 ± 0.16	0.77 ± 0.04
	Ours	$\mathbf{0.74 \pm 0.08}$	$\mathbf{0.62 \pm 0.06}$	$\mathbf{0.78 \pm 0.19}$
Period 3	DOMAIN	0.67 ± 0.19	0.56 ± 0.09	0.71 ± 0.18
	JODIE	0.61 ± 0.08	0.55 ± 0.14	0.59 ± 0.03
	DECENT	0.68 ± 0.12	0.63 ± 0.04	0.71 ± 0.19
	Ours	$\mathbf{0.69 \pm 0.14}$	$\mathbf{0.66 \pm 0.07}$	$\mathbf{0.72 \pm 0.23}$

THE UNIVERSITY OF IOWA

[1] M. Monsalve, S. Pemmaraju, S. Johnson, and P. M. Polgreen, “Improving risk prediction of clostridium difficile infection using temporal event-pairs,” in IEEE ICHI, 2015

MICU Transfer Prediction

Results

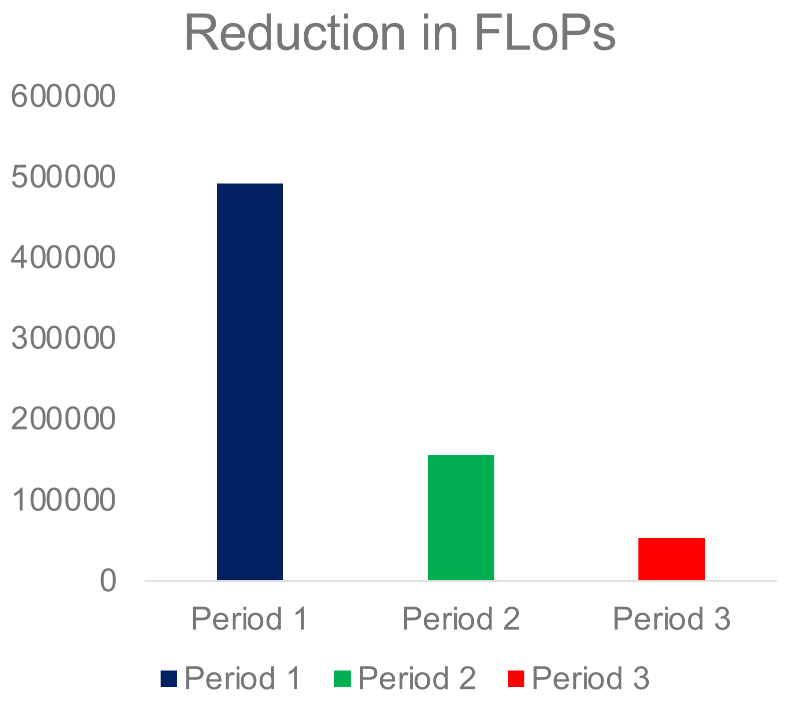
- Forecast whether a patient is at risk of transfer to a Medical Intensive Care Unit (MICU)
- Binary Classification Problem:
  - Instance: Patient at time t and features at that time
  - Label: Binary indicator of MICU transfer in the next day
- Evaluation Metric: ROC-AUC Score
- 3- fold cross validation with 30 repetitions

Period	Method	LR	SVM	RF
Period 1	DOMAIN	0.63 ± 0.20	0.52 ± 0.03	0.86 ± 0.13
	JODIE	0.54 ± 0.15	0.51 ± 0.02	0.66 ± 0.04
	DECENT	0.85 ± 0.07	0.71 ± 0.05	0.83 ± 0.05
	Ours	**0.89 ± 0.05**	**0.77 ± 0.08**	**0.87 ± 0.03**
Period 2	DOMAIN	0.68 ± 0.12	0.57 ± 0.13	0.71 ± 0.07
	JODIE	0.59 ± 0.05	0.52 ± 0.10	0.55 ± 0.01
	DECENT	0.72 ± 0.07	0.65 ± 0.10	0.86 ± 0.03
	Ours	**0.76 ± 0.02**	**0.72 ± 0.03**	**0.89 ± 0.09**
Period 3	DOMAIN	0.67 ± 0.13	0.56 ± 0.02	0.81 ± 0.03
	JODIE	0.61 ± 0.08	0.52 ± 0.18	0.62 ± 0.12
	DECENT	**0.85 ± 0.07**	0.67 ± 0.01	**0.87 ± 0.18**
	Ours	0.84 ± 0.12	**0.71 ± 0.01**	**0.87 ± 0.08**

Empirical Verification of Continual Adaptation

Results

- The model training is very resource intensive
- We used the continual learning formulation to reduce training time without costing too much on the quality of embeddings
- Our model will require more operations to train for the first period. But it will require much less operations to train on the data for the subsequent periods
- We validate this intuition by profiling the proxy of FLoPs (MACs) required to train the model to construct dynamic embeddings across periods
- Note that the number of FLoPs required to pre-train our BERT model is excluded from our analysis



THE UNIVERSITY OF IOWA

Conclusion

- The learned patient embeddings incorporate both the interactions and the clinical notes
- We use continual learning to reduce the time for training incoming heterogenous and dynamic batches of interactions and notes
- We evaluate the performance of the learned embeddings over the predictive tasks:
  - CDI Incidence Prediction
  - MICU Transfer Prediction
- Our proposed model outperforms state-of-the-art baselines across both the tasks
- Our continual learning formulation leads to faster training of model parameters in subsequent batches

THE UNIVERSITY OF IOWA

Team

Akash Choudhuri

Hankyu Jang

Alberto M. Segre

Philip M. Polgreen

Kishlay Jha

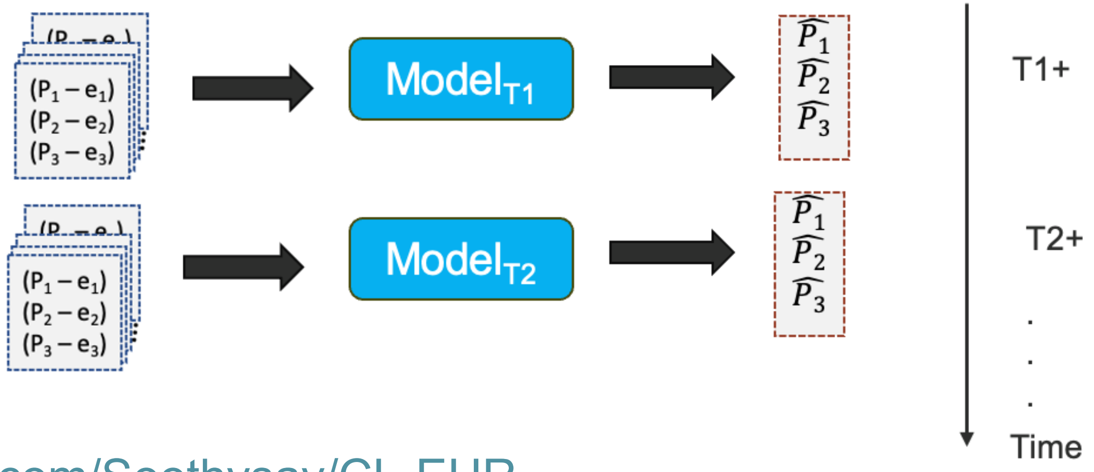
Bijaya Adhikari

Special Thanks

comp|epi computational epidemiology research

CDC

The University of Iowa



Code: https://github.com/Soothysay/CL-EHR

The University of Iowa

Contact:
akash-choudhuri@uiowa.edu