Predicting Early Inpatient Readmission of Psychiatric Patients with Insurance Claims Data

Sooraj Boominathan, Christopher Wang, Kevin Weng

Massachusetts Institute of Technology

6.S897: Machine Learning for Healthcare

Introduction

- Early patient readmission is major, preventable driver of psychiatric healthcare costs
  - Substantial concern for chronic / recurring disorders
- Ability to predict readmission can help develop interventions targeted towards patients at risk and reduce costs
- Prediction from admission data difficult due to lack of clear signal of patient severity / symptoms in diagnosis / procedure codes
- **Goal:** (1) Predict readmission within various time windows, (2) Understand underlying phenotypes that explain readmission
- Previous work has used clinical text to predict readmission [1]
- We use insurance claims data to predict readmission - subset of the information available to physician at time of treatment

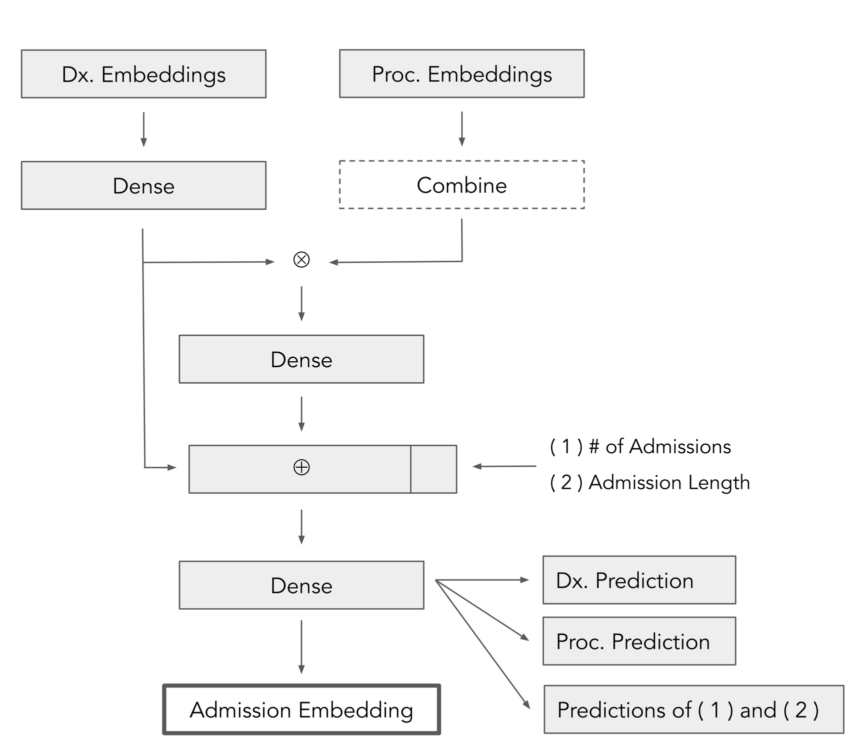
Methods

Feature Construction

- **Basic one-hot representation:** one-hot vectors of a patient's diagnosis, procedure, and medication codes
  - Additional features: Number of past visits, admission duration, time since previous admission
  - Diagnosis / procedure codes in admission history collapsed into same vector as current admission
- **Basic embedding representation:** Used 300-dim. embeddings of medical codes learned from claims data [3]
  - Codes for similar / related concepts close to one another in embedding space
  - Averaged all embeddings associated with admission to obtain embedding for a given admission
- **Labels:** Binary - was patient readmitted within specified time window from admission discharge date?

Learning Admission Representations

- **Goal**: Learn low-dimensional representations of individual inpatient admissions
- Developed an admission "autoencoder" architecture which takes pre-trained embeddings of ICD9 and CPT codes [3] as input
- Train neural network to learn 300-dim. representation of admission that predicts the diagnoses / procedures associated with admission
- Captures interactions between procedures / diagnoses within admissions



Prediction

- Trained logistic regression models and simple feedforward neural networks with dropout
- Used 80/20 train/validation split for hyperparameter tuning

Topic Modeling

- Used Latent Dirichlet Allocation to model the underlying health states as topics
- "Document"= patient admission history, "Words" = diagnosis/ procedure/medication codes for patient

Dataset

- Used MarketScan Database, a health insurance claims dataset containing demographics, diagnosis, procedure, and medication codes from 2011-15 [2]
- Considered patients with an inpatient admission anytime between 2011-14 for a mood disorder (ICD code=296.xx)
  - Bipolar disorder, depression, schizophrenia, etc.
- **Cohort**: Inpatient admissions for these patients with a mental disorder diagnosis group and discharge to home
  - Train set: admissions from 2011-13
  - Test set: admissions from 2014 (removed any patient overlap with train set)

Dataset Statistics		
	Train	Test
Admissions	12,768	3,062
Patients	9,417	2,616

Cohort Readmission Rates			
Time Window	Overall	Train (2011-13)	Test (2014)
30 day	0.127	0.132	0.108
90 day	0.213	0.222	0.174
180 day	0.275	0.288	0.223

Results

Prediction

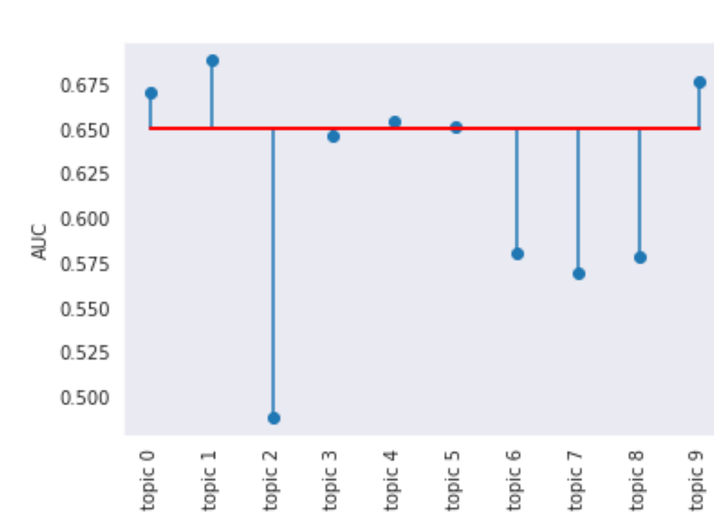
- Embeddings significantly improved test AUC on 30 day prediction task; smaller improvements observed for 90 and 180 day tasks

Task	LR + One Hot	LR + One Hot + History	LR + Embeddings	NN + Embeddings
30 day	0.619	0.621	0.648	**0.655**
90 day	0.629	0.629	0.640	**0.643**
180 day	0.639	0.638	0.644	**0.645**

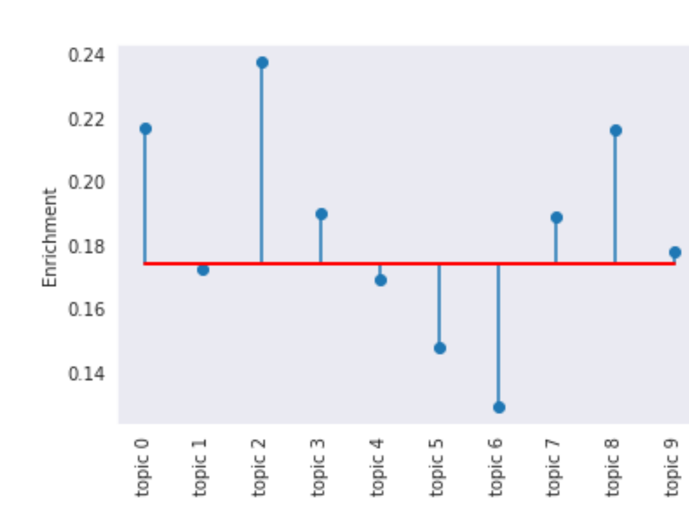
- Neural networks trained on embeddings avoided overfitting with less regularization than using one-hot vectors
- Including past history of diagnoses / procedures or medication data did not significantly help prediction
- Modeling history with LSTMs did not yield improvement
- Top positive predictors for 30-day readmission: # of visits, bipolar disorder diagnosis, schizoaffective disorder, unspecified psychosis, 18-34 age group

Topic Modeling

Topic	Words in order of importance
0	Electrocardiogram report, Unspecified essential hypertension, Age 55-64, Chest x-ray, Chest pain (unspecified), Diabetes (no complications)
2	Abdominal pain (unspecified site), Abdominal procedures, Other chronic pain, Female, Nausea with vomiting, Bipolar disorder (unspecified)
6	Female, Region - South, Suicidal ideation, Major depressive affective disorder (recurrent, severe, no psychotic behavior), Major depressive affective disorder (single episode), Age 35-44



Readmission enrichment across topics



Test AUC across topics

Conclusions / Future Work

- Learned representations for individual admissions improves readmission prediction, particularly in short term
- Modeling topics for patient data reveals diversity of underlying health states
- **Future work:** (1) Understand causal relationships between underlying health states and readmission, (2) Train models fine-tuned for performance on different subpopulations, (3) Model temporal structure of admissions more effectively

References

[1] Rumshisky, A., et al. "Predicting early psychiatric readmission with natural language processing of narrative discharge summaries." Translational psychiatry 6.10 (2016): e921

[2] Adamson, David M., Stella Chang, and Leigh G. Hansen. "Health research data for the real world: the MarketScan databases." *New York: Thompson Healthcare* (2008).

[3] Choi, Youngduck, Chill Yi-I. Chiu, and David Sontag. "Learning low-dimensional representations of medical concepts." *AMIA Summits on Translational Science Proceedings* 2016 (2016): 41