MiME: Multilevel Medical Embedding of Electronic Health Records for Predictive Healthcare

Edward Choi*[1], Cao Xiao[2], Walter F. Stewart[3], Jimeng Sun[4]

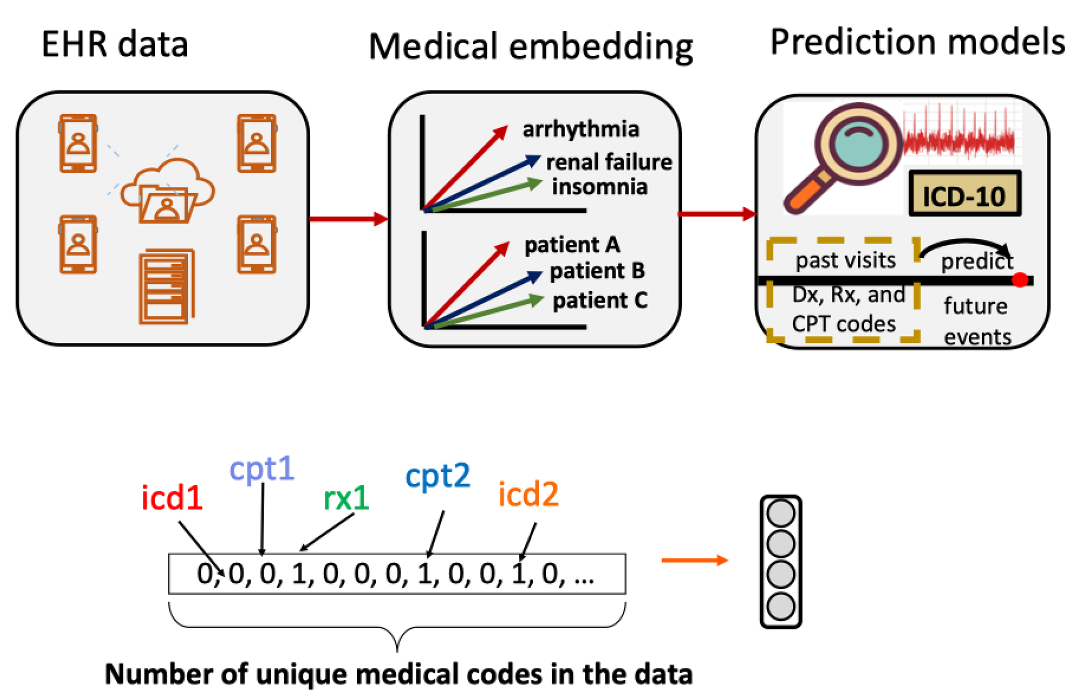
1 Google Brain, 2 IBM Research, 3 Hint Consultants 4 Georgia Institute of Technology * Work done at Georgia Tech

Introduction

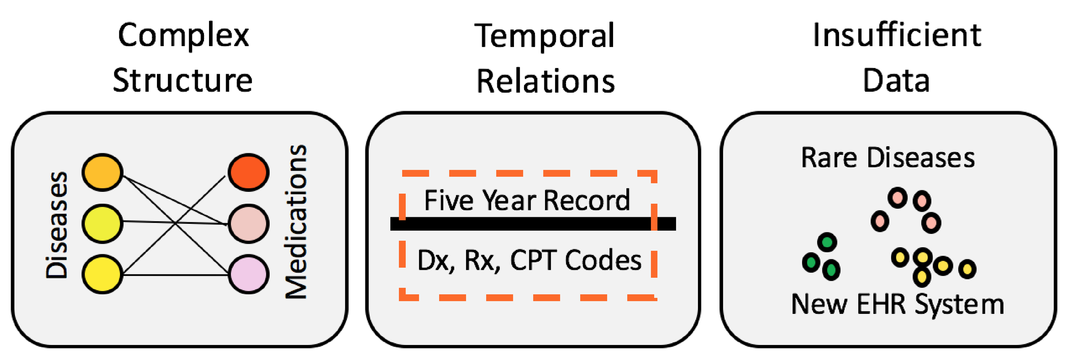
We propose Multilevel Medical Embedding (MiME) which leverage the multilevel structure of electronice health record (EHR) data without the need for external labels. The main ideas of MiME include

1. Modeling the relationships bewteen diagnosis codes and treatment codes can accurately capture the distinct patterns of patient states.
2. Auxiliary tasks of predicting diagnosis and medication within a visit inject the knowledge of EHR data into the embedding process.

Medical Embedding



Medical Embedding takes patient history (represented by medical codes of diagnosis, medication, and procedures) as input, and outputs vectors of patients or medical concepts that can be used in predictive models.



The challenges related to medical embedding includes complex structure within a visit, temporal relations across visits and small datasets.

Related works

	Capture Diagnosis-Treatment Relations	Temporal Relations	Insufficient Data	No Need External Info.
Med2Vec (KDD' 16)		✓		✓
T-LSTM (KDD' 17)		✓		✓
GRAM (KDD' 17)		✓	✓	
DiPole (KDD' 17)		✓		
MiME	✓	✓	✓	✓

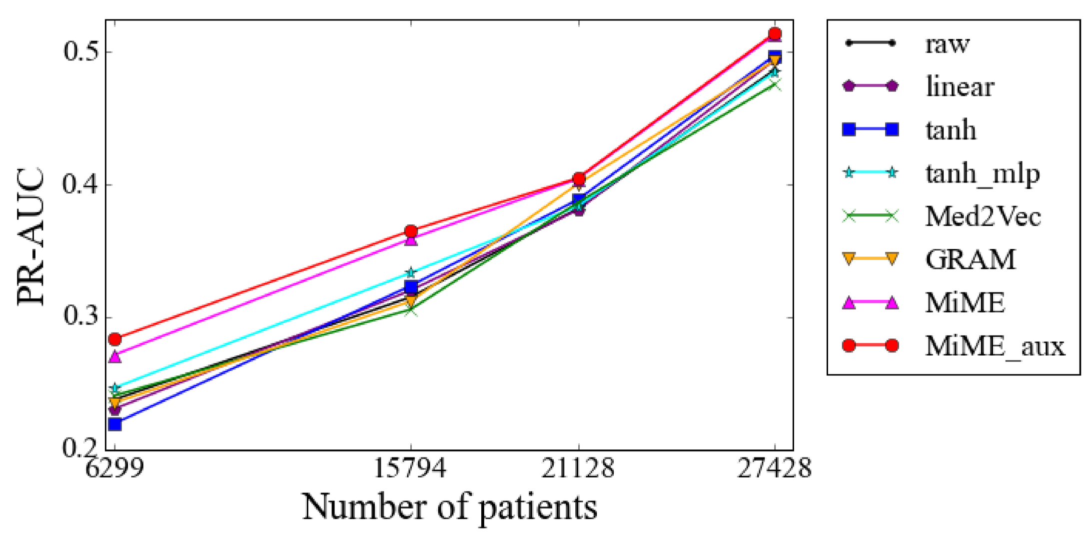
MiME is an unsupervised method that captures diagnosis-treamtent and temporal relations. MiME works well on smaller datasets.

Results

Heart failure prediction: We conducted all our experiments using EHR data provided by Sutter Health. The objective is to predict the onset of heart failure (HF), given a 18-month observation period for each patient.

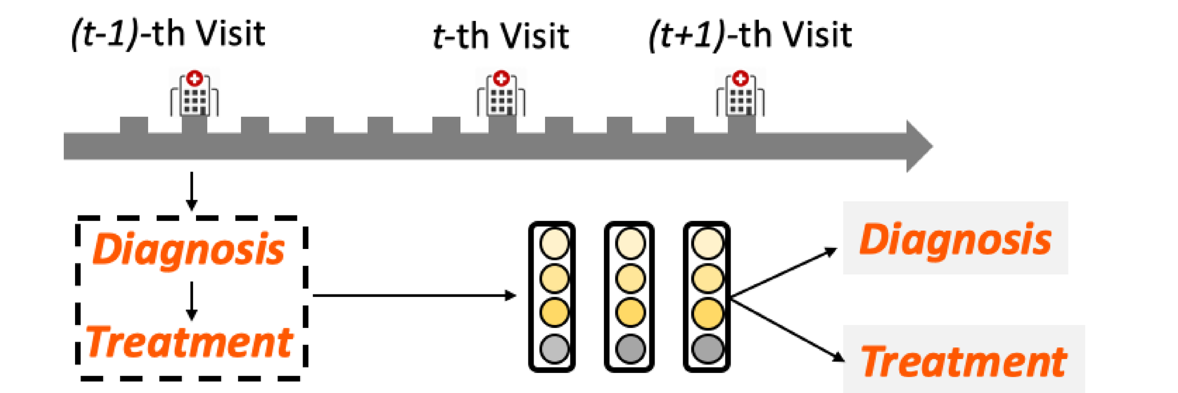
Table 1: Statistics of the dataset.

# of patients	30,764
# of visits	616,073
Avg. # of visits per patient	20.0
# of unique codes	2,311 (Dx:388, Rx:99, Proc:1,824)
Avg. # of Dx per visit	1.93 (Max: 29)
Avg. # of Rx per diagnosis	0.31 (Max: 17)
Avg. # of Proc. per diagnosis	0.36 (Max: 10)



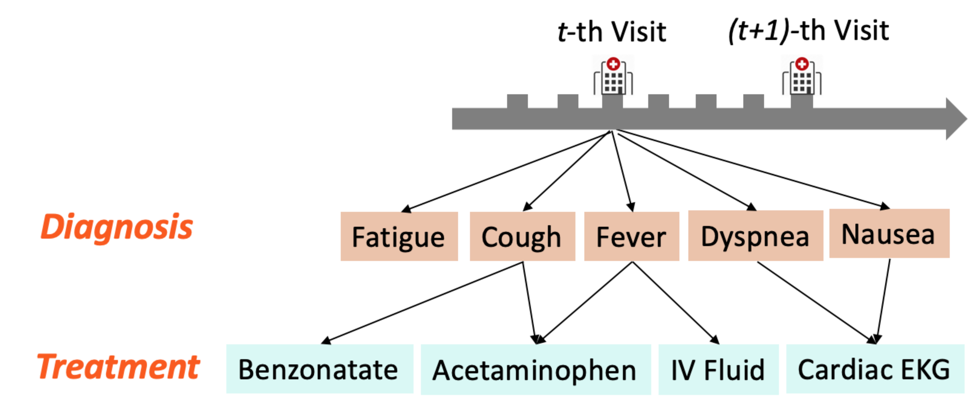
Test PR-AUC of HF prediction when increasing training data size. MiME performs significantly better for the smaller training data. MiME shows superior prediction performance, especially when patient status is complicated (patient receives many medical codes), or insufficient training data.

MiME Architecture



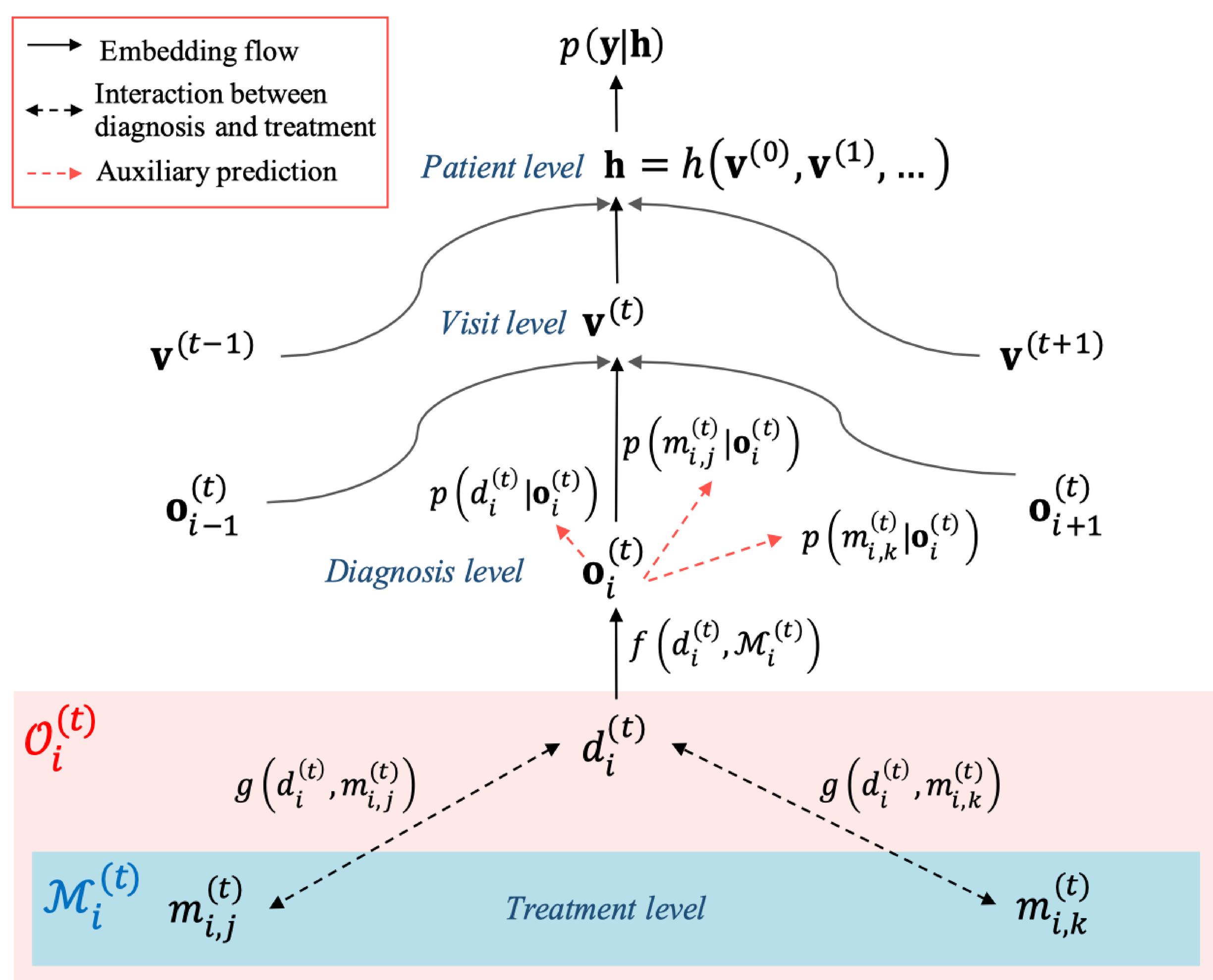
1. Embed **internal** multilevel structure to **capture diagnosis-treatment relations** and model patient states more accurately.

2. Joint training with auxiliary tasks to **help insufficient data.**



A good medical embedding method needs to capture 1) longitudinal/temporal relationship across multiple visits over time and 2) diagnosis-treatment relationship within a visit.

MiME Model Details



MiME is an unsupervised method that captures diagnosis-treamtent and temporal relations. MiME works well on smaller datasets. MiME explicitly captures the hierarchy between diagnosis codes and treatment codes. A single visit $\mathcal{V}$ (omitting the index (t)) consists of two levels: the diagnosis level, and the treatment level.
The diagram illustrates how MiME builds the representation of a visit in a bottom-up fashion via multilevel embedding.

Notations: In a single diagnosis object $\mathcal{O}_i$, a diagnosis code d_i and its associated treatment codes $\mathcal{M}_i$ are used to obtain a vector representation of $\mathcal{O}_i$, $\mathbf{o}_i$. Then multiple diagnosis object embeddings $\mathbf{o}_0, \dots, \mathbf{o}_{|\mathcal{V}|}$ in a single visit are used to obtain a visit embedding $\mathbf{v}$, which in turn forms a patient embedding $\mathbf{h}$ with other visit embeddings.

1. Capture the interaction between i-th diagnosis d_i and its j-th treatment $m_{i,j}$.
   - $g(d_i, m_{i,j}) = \sigma(\mathbf{W}_m r(d_i)) \odot r(m_{i,j})$
   - $r(\cdot)$: a helper function to retrieve the embedding vector of d_i or $m_{i,j}$.
2. Generate representation vector $\mathbf{o}_i$ for i-th diagnosis.
   - $f(d_i, \mathcal{M}_i) = \mathbf{o}_i = \sigma\Big(\mathbf{W}_o\Big(\underbrace{r(d_i) + \sum_{j}^{|\mathcal{M}_i|} g(d_i, m_{i,j})}_{\text{G: used for skip-connection}}\Big)\Big) + G$
3. Generate representation vector $\mathbf{v}$ for (t-th) visit.
   - $\mathbf{v} = \sigma\Big(\mathbf{W}_v\Big(\underbrace{\sum_{i}^{|\mathcal{V}|} f(d_i, \mathcal{M}_i)}_{\text{F: used for skip-connection}}\Big)\Big) + F$
4. Generate patient representation vector $\mathbf{h}$.
   - Function $h(\mathbf{v}^{(0)}, \mathbf{v}^{(1)}, \dots)$ could be *sum*, RNN, CNN, Transformer, etc.
5. Auxiliary prediction task for $\mathbf{o}_i$
   - An effective embedding $\mathbf{o}_i$ should be able to predict what it represents; diagnosis d_i and treatment $m_{i,j}$.
   - $\hat{d}_i^{(t)} = p(d_i^{(t)}|\mathbf{o}_i^{(t)}) = \text{softmax}(\mathbf{U}_d \mathbf{o}_i^{(t)})$
   - $\hat{m}_{i,j}^{(t)} = p(m_{i,j}^{(t)}|\mathbf{o}_i^{(t)}) = \sigma(\mathbf{U}_m \mathbf{o}_i^{(t)})$

Contact and Code

Contact: jsun@cc.gatech.edu

Paper: https://arxiv.org/abs/1810.09593 **Code:** https://github.com/mp2893/mime

References

[1] Edward Choi, Cao Xiao, Walter F Stewart, and Jimeng Sun. Mime: Multilevel embedding of electronic health records for predictive healthcare. In *Neural Information Processing Systems*, 2018.