Multi-layer Representation Learning for Medical Concepts

Speaker: Shih-Han Lo

Advisor: Professor Jia-Ling Koh

Author: Edward Choi, Mohammad Taha Bahadori, Elizabeth Searles, Catherine Coffey, Michael Thompson, James Bost, Javier Tajedor-Sojo, Jimeng Sun

Date: 2017/10/31

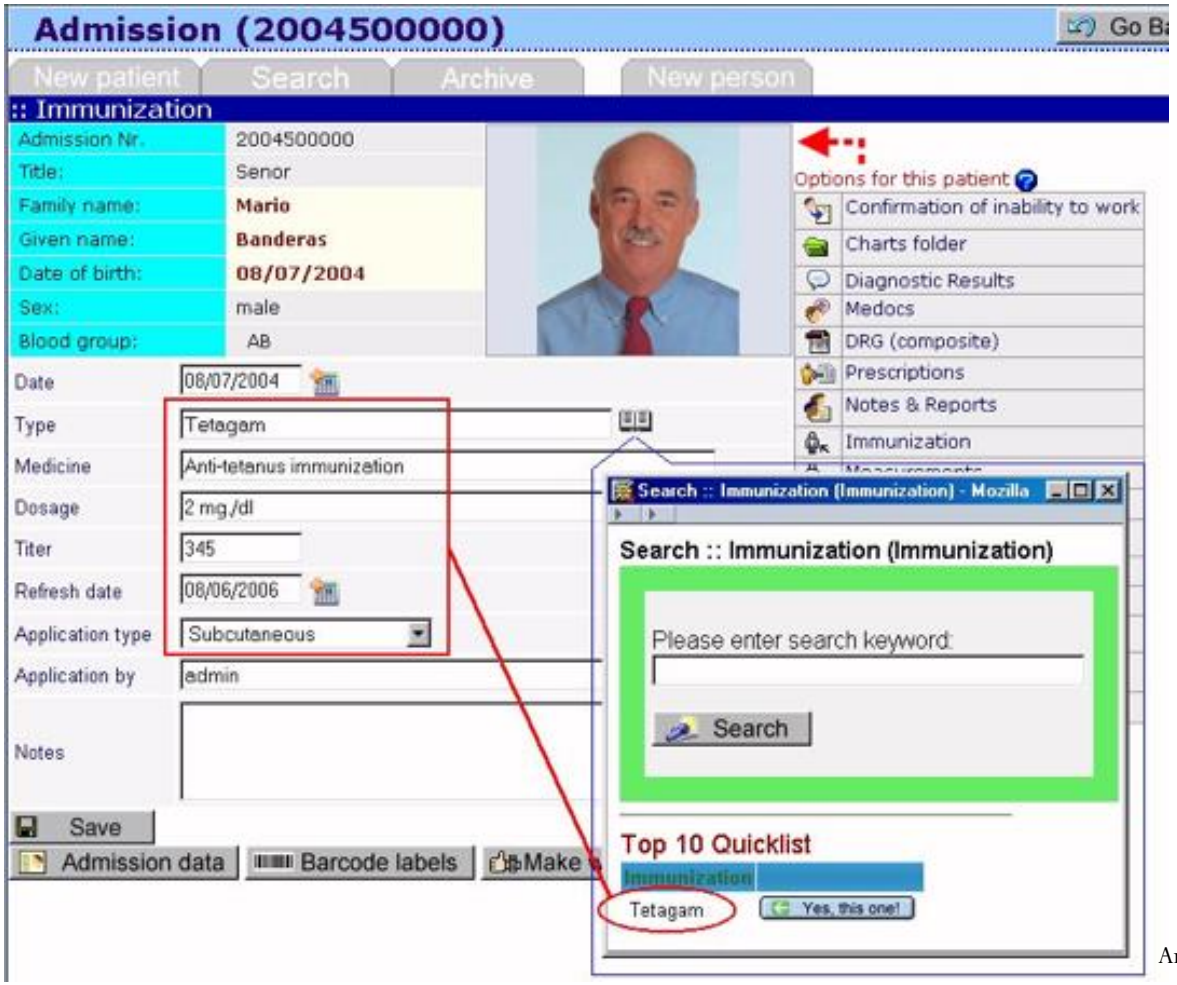
Source: KDD ’16

Outline

- Introduction
- Method
- Experiment
- Conclusion

Introduction

- **Motivation**



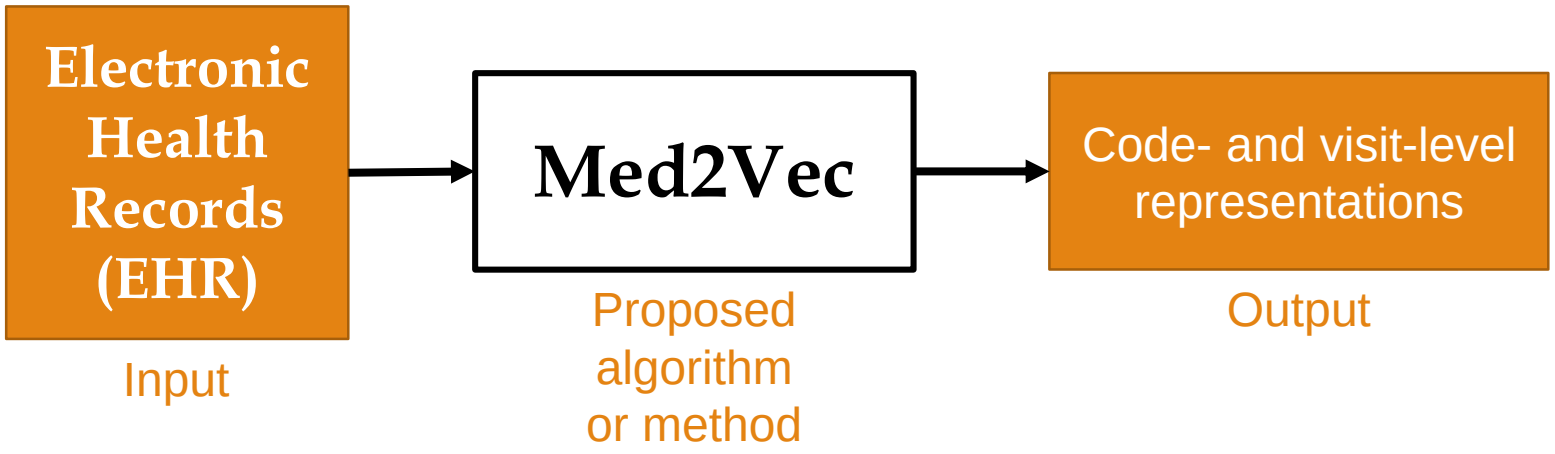
Ars Technica

Introduction

- **Purpose**
  - Learn interpretable representations.
  - Enable clinical applications to offer more than just improved performances.

Introduction

- **Framework**



Introduction

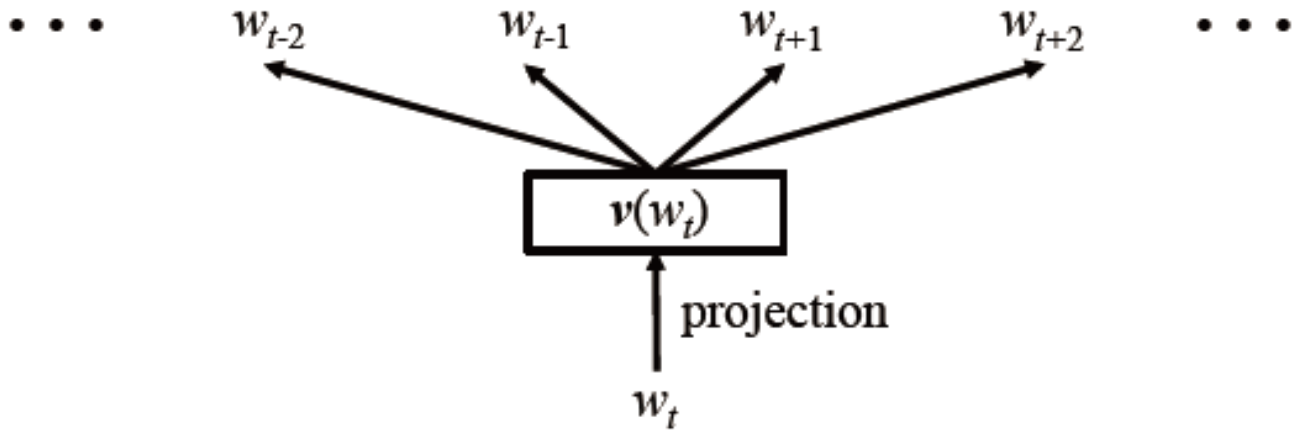


Figure 1: Skip-gram model architecture: $v(w_t)$ is a vector representation for the word w_t. The goal of Skip-gram is to learn vector representations of words that are good at predicting neighboring words.

Introduction

- **EHR structure**
  - The set of all medical codes: $c_1, c_2, \ldots, c_{|\mathcal{C}|}$
  - Sequence of visits: $V_1, \ldots, V_T$ where $V_t \subseteq \mathcal{C}$.

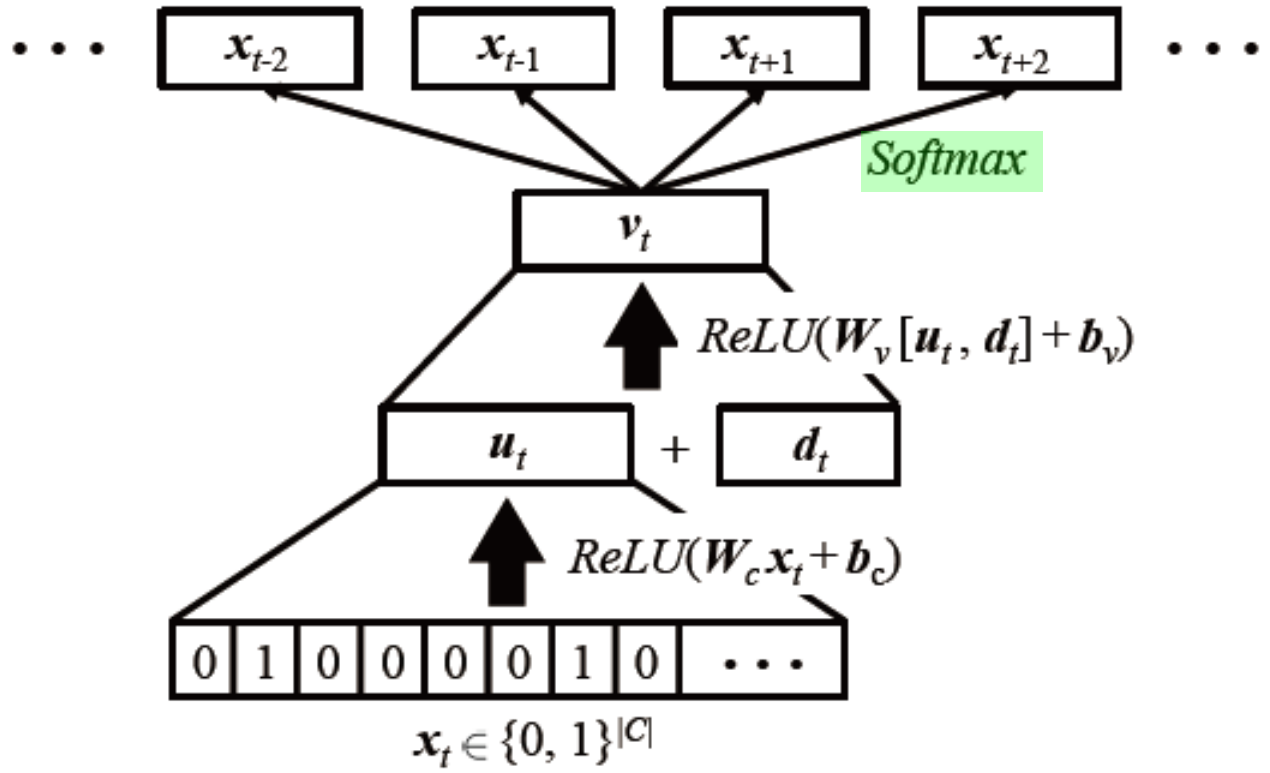
  - The goal of **Med2Vec** is to learn two types of representations:
    - Code representations
    - Visit representations

Outline

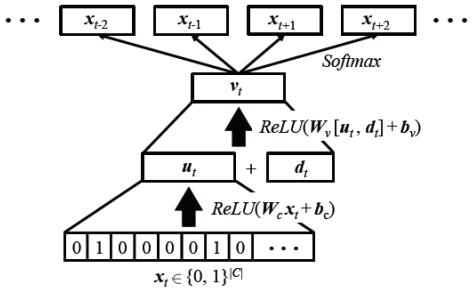
- Introduction
- Method
- Experiment
- Conclusion

Method

- **Med2Vec** architecture



Method

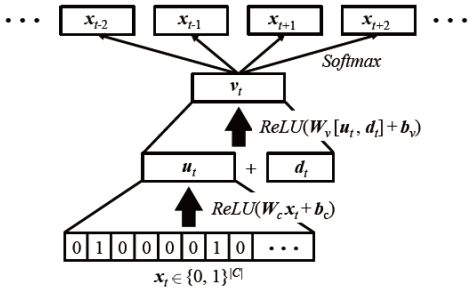


- **Learning from the visit-level representation**
  - We minimize the cross entropy error as follows:

$$\min_{\boldsymbol{W}_s, \boldsymbol{b}_s} \frac{1}{T} \sum_{t=1}^{T} \sum_{-w \leq i \leq w, i \neq 0} -\boldsymbol{x}_{t+i}{}^{\top} \log \hat{\boldsymbol{y}}_t - (\mathbf{1} - \boldsymbol{x}_{t+i})^{\top} \log(\mathbf{1} - \hat{\boldsymbol{y}}_t) \tag{2}$$

$$\text{where} \qquad \hat{\boldsymbol{y}}_t = \frac{\exp(\boldsymbol{W}_s \boldsymbol{v}_t + \boldsymbol{b}_s)}{\sum_{j=1}^{|C|} \exp(\boldsymbol{W}_s[j,:]\boldsymbol{v}_t + \boldsymbol{b}_s[j])}$$

Method

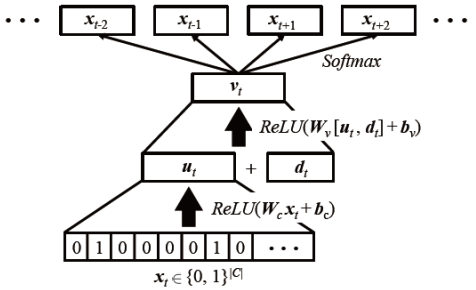


- **Learning from the code-level representation**
  - The code-level representation can be learned by maximizing the following likelihood.

$$\min_{\boldsymbol{W}'_c} -\frac{1}{T}\sum_{t=1}^{T}\sum_{i:c_i\in V_t}\sum_{j:c_j\in V_t, j\neq i}\log p(c_j|c_i), \quad (3)$$

$$\text{where} \quad p(c_j|c_i) = \frac{\exp\left(\boldsymbol{W}'_c[:,j]^\top \boldsymbol{W}'_c[:,i]\right)}{\sum_{k=1}^{|\mathcal{C}|}\exp\left(\boldsymbol{W}'_c[:,k]^\top \boldsymbol{W}'_c[:,i]\right)}. \quad (4)$$

Method



- **Unified training**

Function (3)

$$\underset{\boldsymbol{W}_{c,v,s},\boldsymbol{b}_{c,v,s}}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1}^{T}\Bigg\{ -\sum_{i:c_i\in V_t}\sum_{j:c_j\in V_t, j\neq i}\log p(c_j|c_i) + \sum_{-w\leq k\leq w, k\neq 0} -\boldsymbol{x}_{t+k}^{\top}\log\hat{\boldsymbol{y}}_t - (1-\boldsymbol{x}_{t+k})^{\top}\log(1-\hat{\boldsymbol{y}}_t)\Bigg\}$$

Function (2)

Method

- **Interpretation of learned representations**
  - Code representations
    - Non-negative matrix factorization (NMF)
    - $\text{argsort}(\boldsymbol{W}_c[i,:])[1:k]$
  - Visit representations
    - $\text{argsort}(\boldsymbol{W}_v[i,:])[1:k]$

Method

Coordinate 112	Coordinate 152	Coordinate 141
Kidney replaced by transplant (V42.0)	X-ray, knee (P)	Cystic fibrosis (277.02)
Hb-SS disease without crisis (282.61)	X-ray, thoracolumbar (P)	Intracranial injury (854.00)
Heart replaced by transplant (V42.1)	Accidents in public building (E849.6)	Persistent mental disorders (294.9)
RBC antibody screening (P)	Activities involving gymnastics (E005.2)	Subdural hemorrhage (432.1)
Complications of transplanted	Struck by objects/persons in sports (E917.0)	Neurofibromatosis (237.71)
bone marrow (996.85)	Encounter for removal of sutures (V58.32)	Other conditions of brain (348.89)
Sickle-cell disease (282.60)	Struck by object in sports (E917.5)	Conductive hearing loss (389.05)
Liver replaced by transplant (V42.7)	Unspecified fracture of ankle (824.8)	Unspecified causes of encephalitis,
Hb-SS disease with crisis (282.62)	Accidents occurring in place for	myelitis, encephalomyelitis (323.9)
Prograf PO (R)	recreation and sport (E849.4)	Sensorineural hearing loss (389.15)
Complications of transplanted heart (996.83)	Activities involving basketball (E007.6)	Intracerebral hemorrhage (431)
Coordinate 184	**Coordinate 190**	**Coordinate 199**
	Down's syndrome (758.0)	
Pain in joint, shoulder region (719.41)	Congenital anomalies (759.89)	Infantile cerebral palsy (343.9)
Pain in joint, lower leg (719.46)	Tuberous sclerosis (759.5)	Congenital quadriplegia (343.2)
Pain in joint, ankle and foot (719.47)	Anomalies of larynx, trachea,	Congenital diplegia (343.0)
Pain in joint, multiple sites (719.49)	and bronchus (748.3)	Quadriplegia (344.00)
Generalized convulsive epilepsy (345.10)	Autosomal deletions (758.39)	Congenital hemiplegia (343.1)
Pain in joint, upper arm (719.42)	Conditions due to anomaly of unspecified	Baclofen 10mg tablet (R)
Cerebral artery occlusion (434.91)	chromosome (758.9)	Wheelchair management (P)
MRI, brain (780.59)	Acquired hypothyroidism (244.9)	Tracheostomy status (V44.0)
Other joint derangement (718.81)	Conditions due to chromosome anomalies (758.89)	Paraplegia (344.1)
Fecal occult blood (790.6)	Anomalies of spleen (759.0)	Baclofen 5mg/ml liquid (R)
	Conditions due to autosomal anomalies (758.5)	

Outline

- Introduction
- Method
- Experiment
- Conclusion

Experiment

- **Datasets**

Dataset	CHOA	CMS
# of patients	550,339	831,210
# of visits	3,359,240	5,464,950
Avg. # of visits per patient	6.1	6.57
# of unique medical codes	28,840	21,033
- # of unique diagnosis codes	10,414	14,111
- # of unique medication codes	12,892	N/A
- # of unique procedure codes	5,534	6,922
Avg. # of codes per visit	7.88	3.19
Max # of codes per visit	440	44
(95%, 99%) percentile # of codes per visit	(22, 53)	(9, 13)

Experiment

- **Evaluation strategies**
  - Code representations
    - Qualitative evaluation by medical experts
    - Quantitative evaluation with baselines: NMI
  - Visit representation
    - Predicting future medical codes
    - Predicting CRG level
    - Baselines: One-hot+, SA, Skip-gram+, GloVe+

Experiment

- **Results**

Table 2: Average score of the medical codes from the relatedness test. 2 was assigned for *related*, 1 for *possible* and 0 for *unrelated*

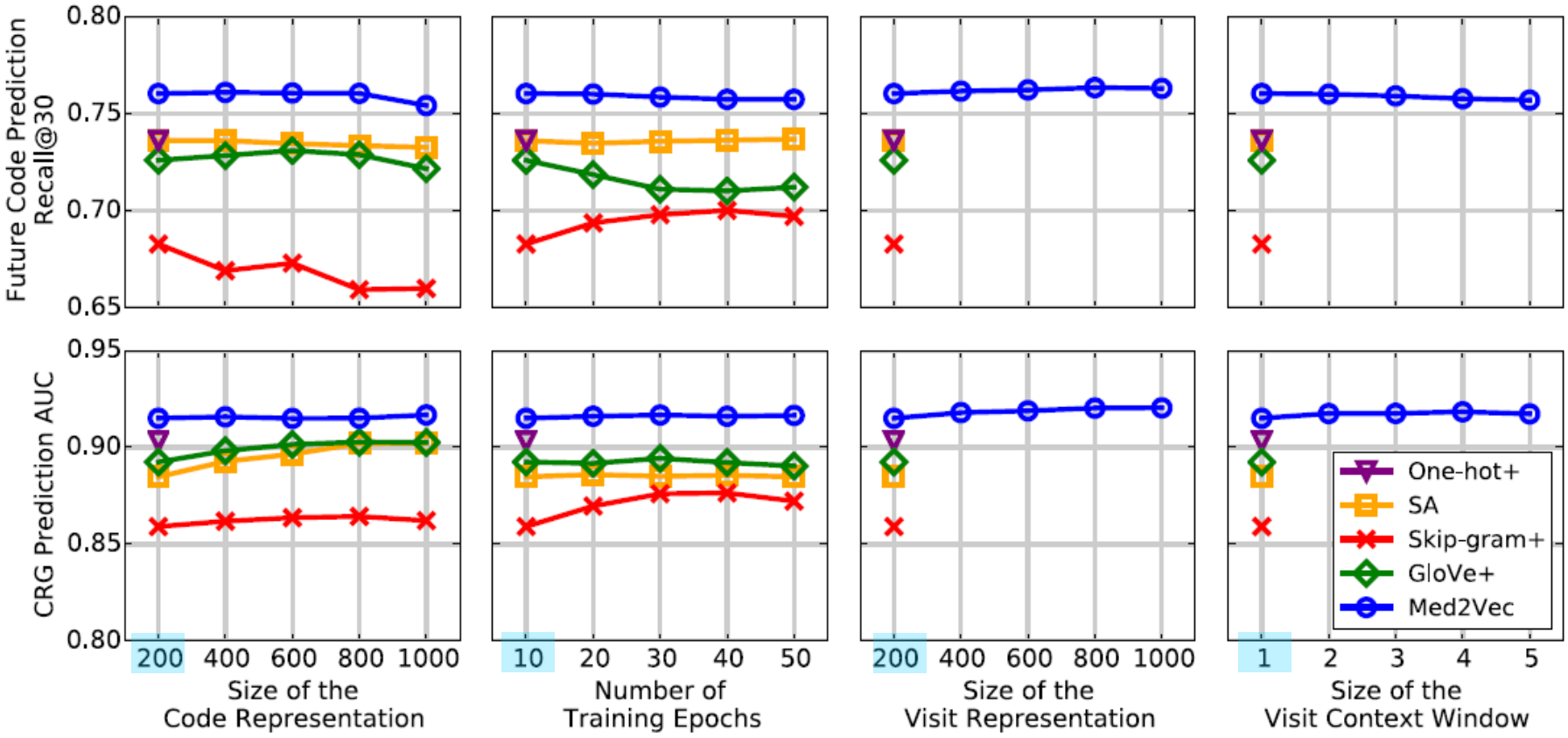
Average	Diagnosis	Medication	Procedure
1.34	1.59	0.95	1.47

Table 3: Clustering NMI of the diagnosis, medication and procedure code representations of various models. All models learned 200 dimensional code vectors. All models except SVD were trained for 10 epochs.

Model	Diagnosis	Medication	Procedure
SVD ($\sigma \boldsymbol{V}^\top$)	0.1824	0.0843	0.1781
Skip-gram	0.2251	0.1216	0.2432
GloVe	0.4205	0.2163	0.3499
Med2Vec	0.2328	0.1089	0.21

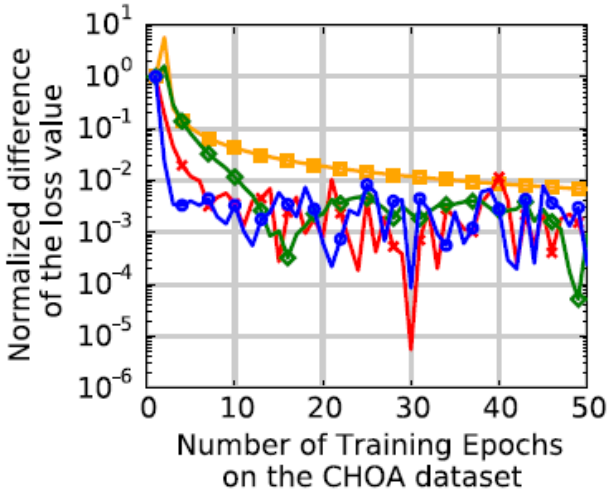
Experiment

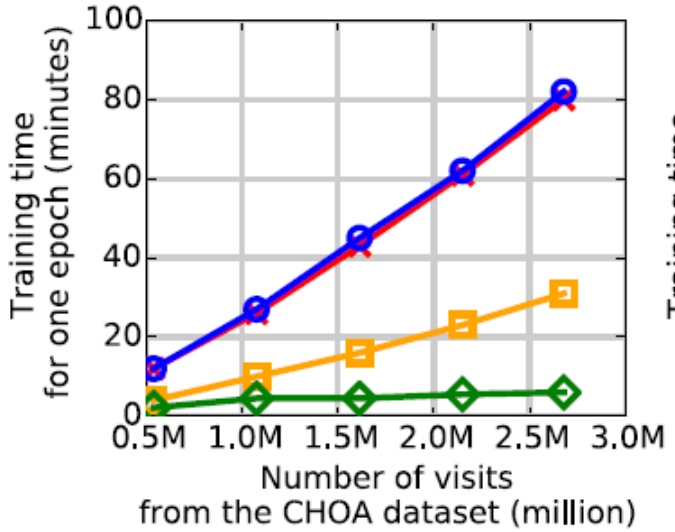
- **Results**

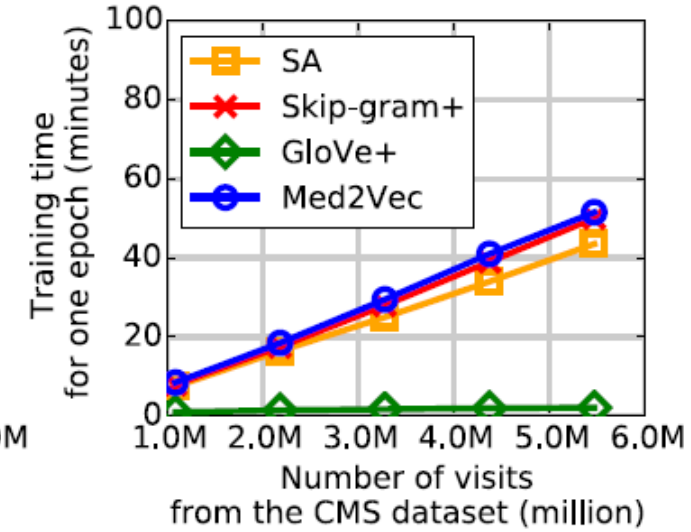


Experiment

- **Results**







Experiment

- **Results**

Table 4: Performance comparison of two Med2Vec models. The top row was trained with the grouped code as mentioned in section 4.4. The bottom row was trained without using the groupers. Both models were trained for 10 epochs with $m, n = 200$, $w = 1$.

Model	Future code prediction	CRG prediction
Grouped codes	0.7605	0.9150
Exact codes	0.7574	0.9155

Outline

- Introduction
- Method
- Experiment
- Conclusion

Conclusion

- We proposed **Med2Vec** for learning lower dimensional representations for medical concepts.
- **Med2Vec** incorporates both code co-occurrence information and visit sequence information of the EHR data.