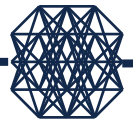
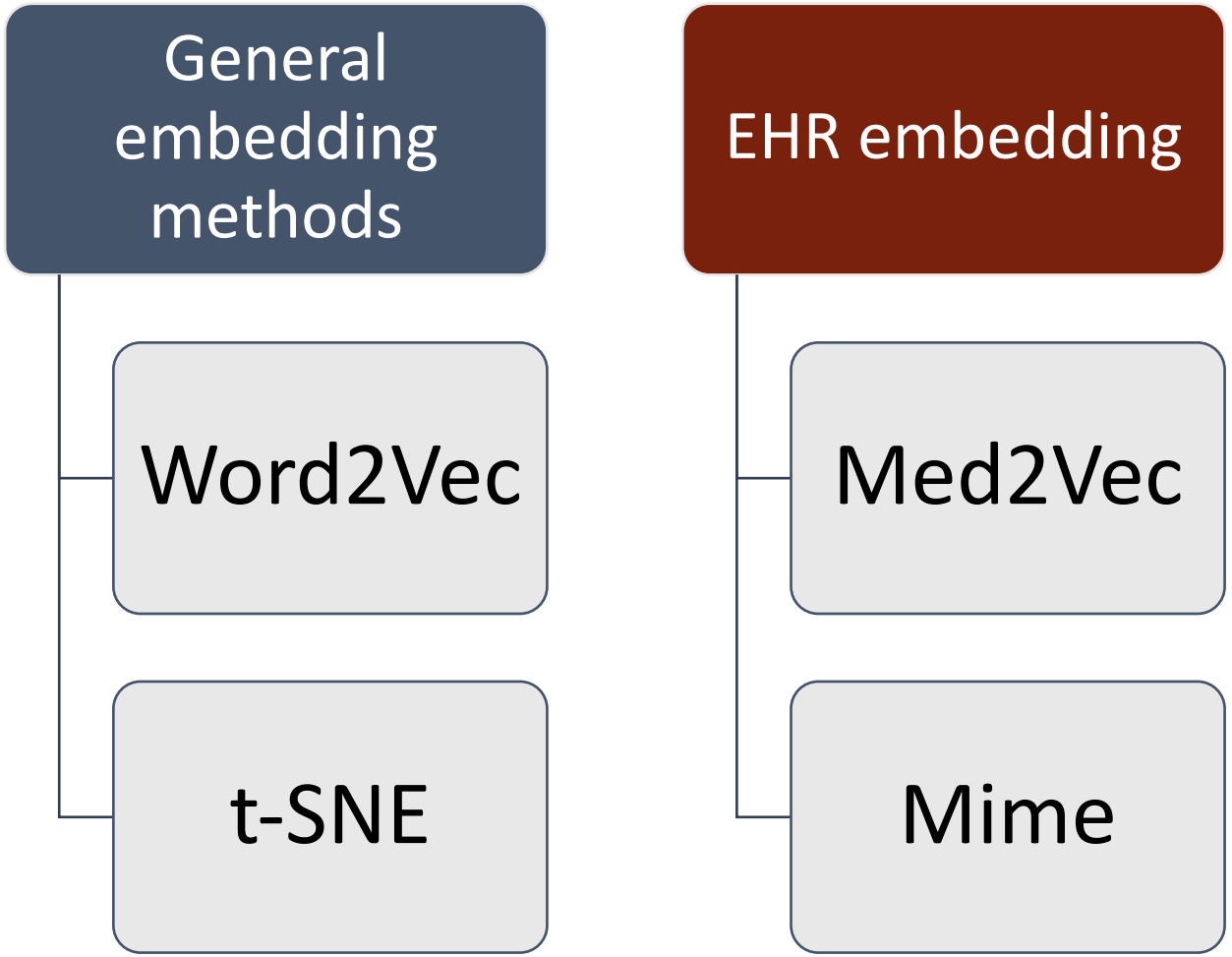
Lecture 5: Embedding

Jimeng Sun

ILLINOIS

Outline





ILLINOIS

Word2Vec

General method

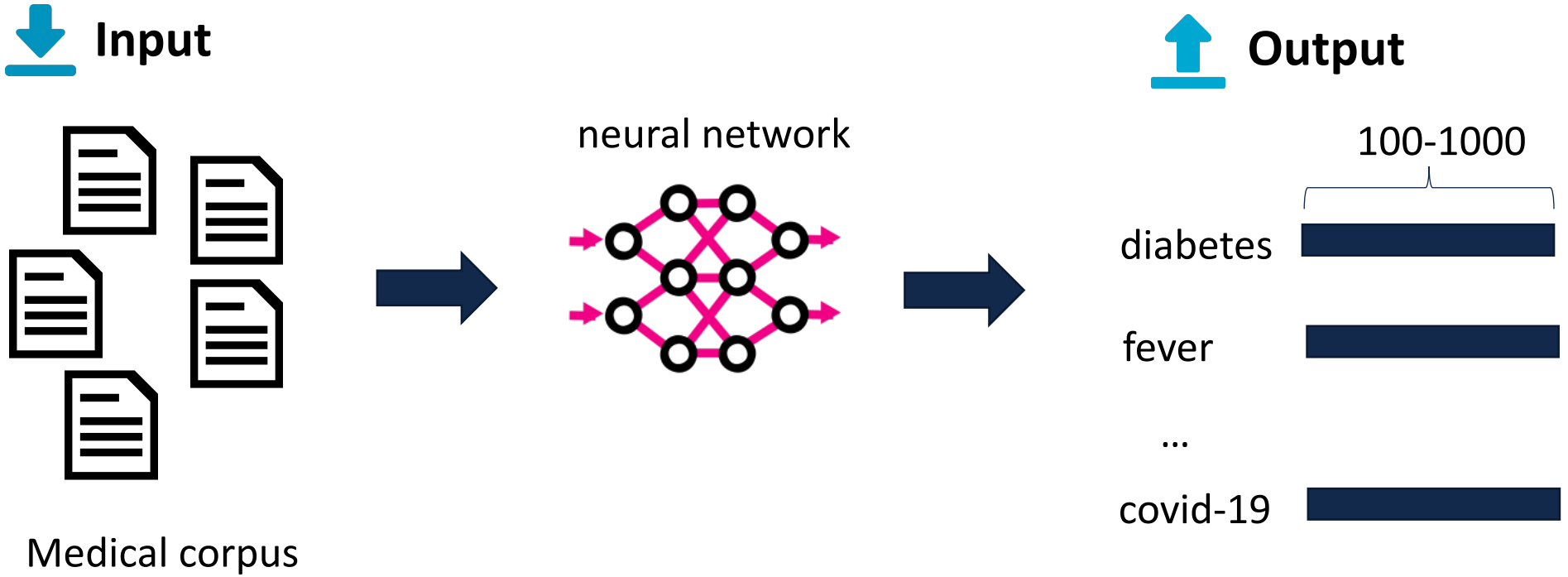
Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. “Distributed Representations of Words and Phrases and Their Compositionality.” *NIPS*

Medical application of word2vec

Choi, E., A. Schuetz, W. F. Stewart, and J. Sun. 2016. “Medical Concept Representation Learning from Electronic Health Records and Its Application on Heart Failure Prediction.” arXiv Preprint arXiv:1602.03686. http://arxiv.org/abs/1602.03686.

Word2Vec

- A neural network to produce vector representations for words



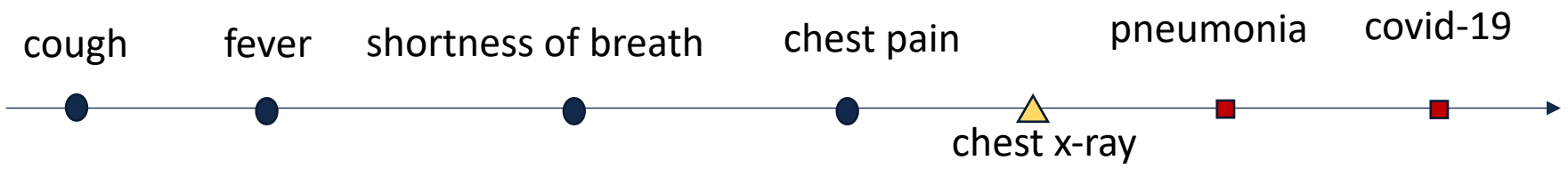
ILLINOIS

Input of Word2Vec

- One-hot encoding vector for each “word”

cough	1	0	0	0	0	0	0	0
fever	0	1	0	0	0	0	0	0
headache	0	0	1	0	0	0	0	0
…								
covid-19	0	0	0	0	0	0	0	1

- Sequence of one-hot vectors for each patient



ILLINOIS

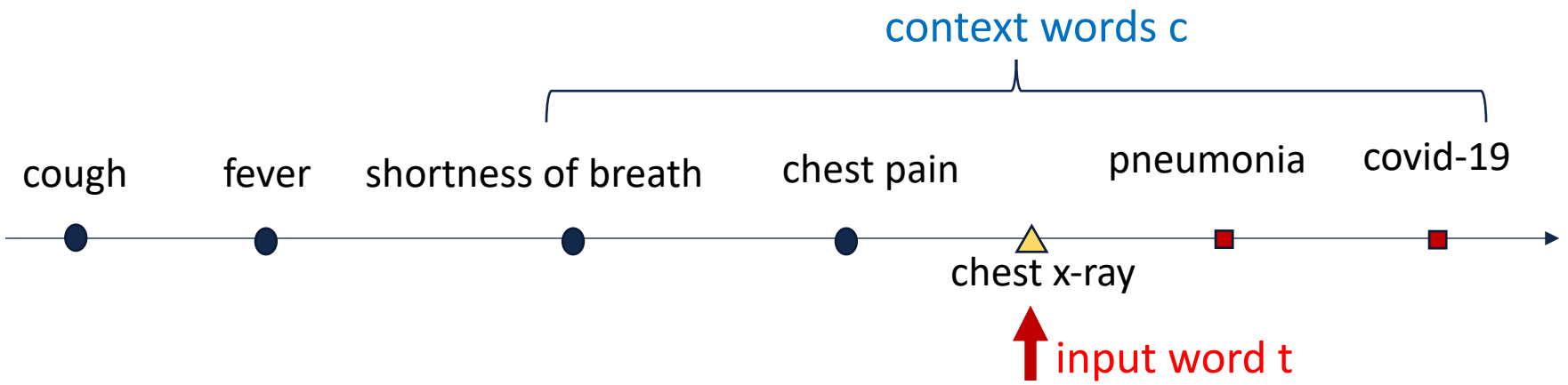
Issue of one-hot vectors

- All words have the same distance to each other

- No distinction between similar words and dissimilar words
  - D (obesity, pneumonia) = D(pneumonia, bronchitis)

- Goal: Find embedding vectors such that similar words are close together and dissimilar words are far apart

Word2Vec formulation: Skip gram

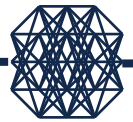


- Maximize $\sum_{\text{all pairs}(c,t)} p(c|t)$
- $p(c|t) = \exp(v_c^T v_t) / \sum_{\text{all words }(w)} \exp(v_w^T v_t)$

☹ Expensive to compute

ILLINOIS

Negative sampling to speed up Word2Vec



- Reformulate the objective to

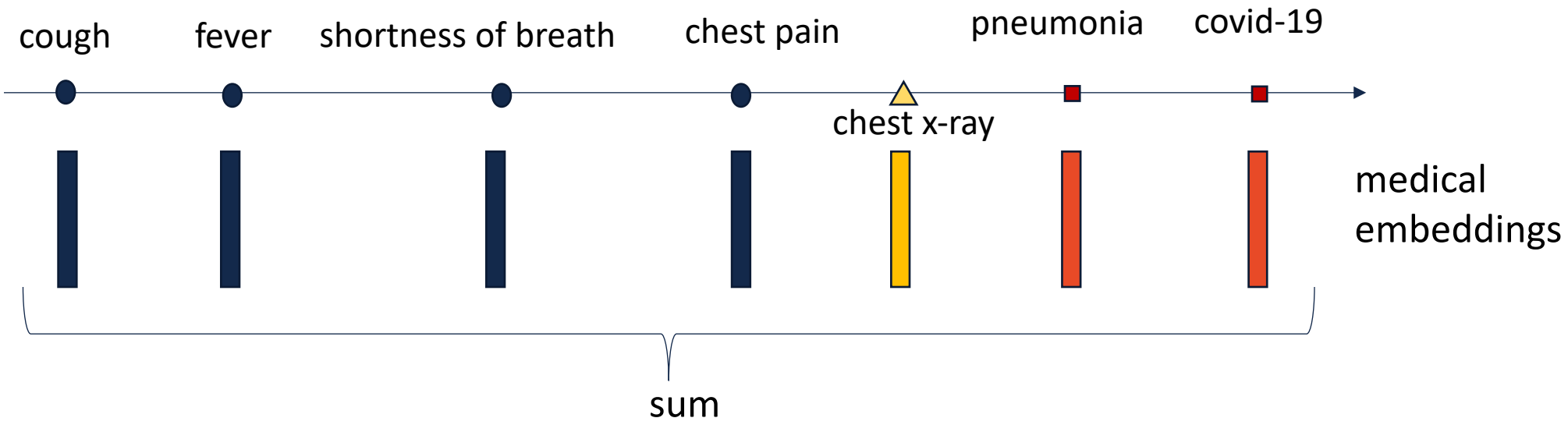
Negative samples

$$\arg\max_{\theta} \underbrace{\prod_{(w,c)\in D} p(D=1|c,w;\theta)}_{\text{(context, input) pairs in the corpus}} \underbrace{\prod_{(w,c)\in D'} p(D=0|c,w;\theta)}_{\text{(random word, input) pairs NOT in the corpus}}$$

$$= \arg\max_{\theta} \sum_{(w,c)\in D} \log \frac{1}{1+e^{-v_c \cdot v_w}} + \sum_{(w,c)\in D'} \log\left(1 - \frac{1}{1+e^{-v_c \cdot v_w}}\right)$$

ILLINOIS

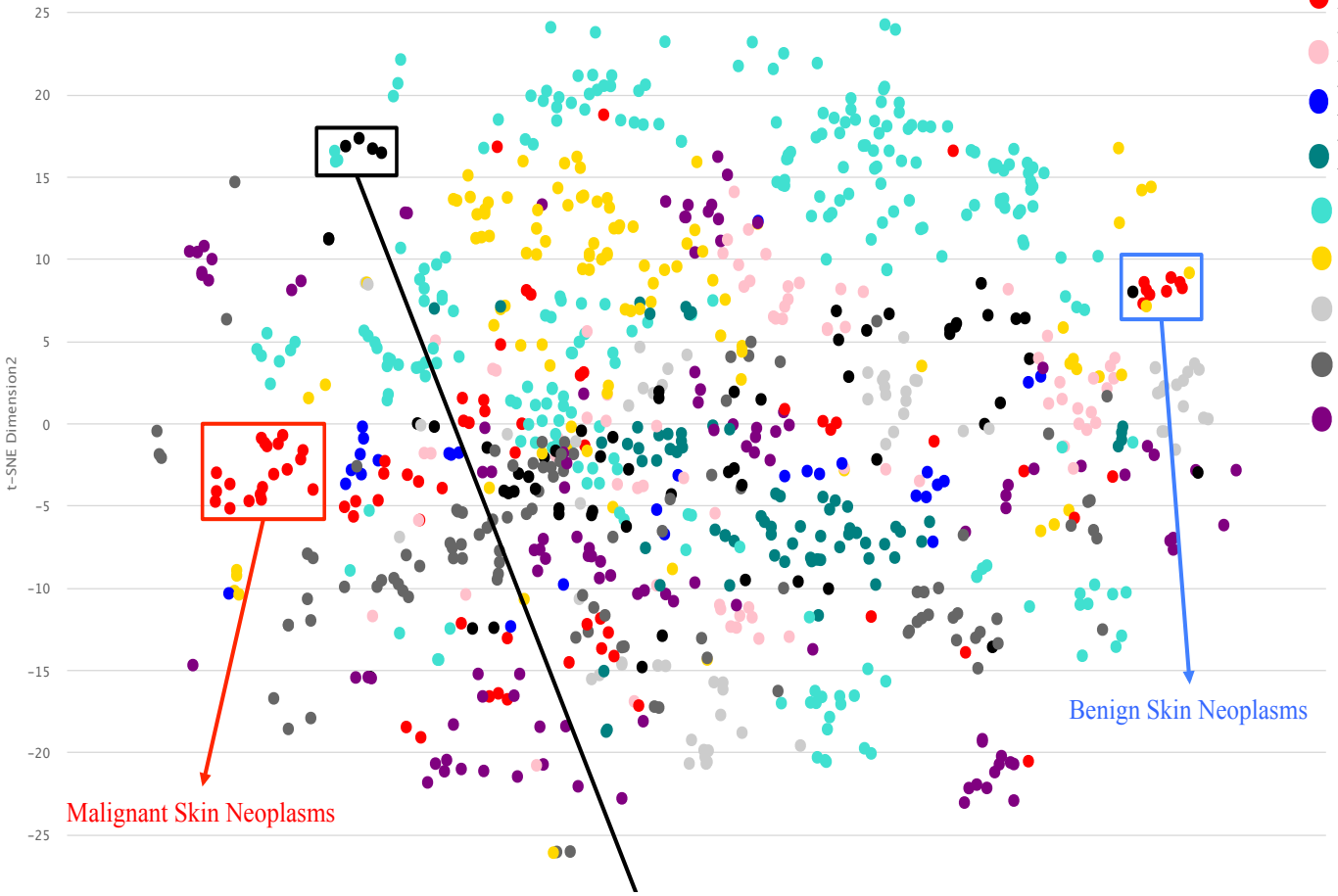
Compute patient representation



- Patient representation = sum of medical embeddings
  - medical embeddings are output of word2vec

ILLINOIS

Visualize word2vec embedding



- Infectious And Parasitic Diseases
- Neoplasms
- Endocrine, Nutritional And Metabolic Diseases, And Immunity Disorders
- Diseases Of The Blood And Blood-Forming Organs
- Mental Disorders
- Diseases Of The Nervous System And Sense Organs
- Diseases Of The Circulatory System
- Diseases Of The Respiratory System
- Diseases Of The Digestive System
- Diseases Of The Genitourinary System

ILLINOIS

Similarity search based on word2vec

	Diagnoses	Medications	Procedures
Acute upper respiratory infections (465.9)	-Bronchitis, not specified as acute or chronic (490) -Cough (786.2) -Acute sinusitis, unspecified (461.9) -Acute bronchitis (466.0) -Acute pharyngitis (462)	-Azithromycin 250 mg po tabs -Promethazine-Codeine 6.25-10 mg/5ml po syrp -Amoxicillin 500 mg po caps -Fluticasone Propionate 50 mcg/act na susp -Flonase 50 mcg/act na susp	-Pulse oximetry single -Serv prov during reg sched eve/wkend/hol hrs -Chest PA & lateral -Gyn cytology (pap) pa -Influenza vac (flu clinic only) 3+yo pa

ILLINOIS

Similarity search based on word2vec

	Diagnoses	Medications	Procedures
Diabetes mellitus (250.02)	-Diabetes mellitus (250.00) -Mixed hyperlipidemia (272.2) -Other abnormal glucose (790.29) -Obesity, unspecified (278.00) -Pure hypercholesterolemia (272.0)	-Metformin hcl 500 mg po tabs -Metformin hcl 1000 mg po tabs -Glucose blood vi strp -Lisinopril 10 mg po tabs -Lisinopril 20 mg po tabs	-Diabetic eye exam (no bill) -Diabetes education, int -Ophthalmology, int -Diabetic foot exam (no bill) -Influenza vac 3+yr (v04.81) im

ILLINOIS

Similarity search based on word2vec

	Diagnoses	Medications	Procedures
Edema (782.3)	-Anemia, unspecified (285.9) -Congestive heart failure, unspecified (428.0) -Unspecified essential hypertension (401.9) -Atrial fibrillation (427.31) -Chronic kidney disease, Stage III (moderate) (585.3)	-Furosemide 20 mg po tabs -Hydrochlorothiazide 25 mg po tabs -Hydrocodone-Acetaminophen 5-500 mg po tabs -Cephalexin 500 mg po caps -Furosemide 40 mg po tabs	-Debridement of nails, 6 or more -OV est pt min serv -EKG -ECG and interpretation -Chest PA & lateral

ILLINOIS

Algebraic operation on word2vec embeddings

	Diagnoses	Medications	Procedures
Hypertension (401.9) **+** **Obesity (278.0)**	-Hyperlipidemia (272.4) -Diabetes (250.00) -Coronary atherosclerosis (414.00) -Hypertension (401.1) -Chronic kidney disease (585.3)	-Hydrochlorothiazide -Valsartan -Nifedipine -Lisinopril -Losartan potassium	N/A

ILLINOIS

Algebraic operation on word2vec embeddings

	Diagnoses	Medications	Procedures
Fever (780.60) + Cough (786.2)	-Pneumonia (486) -Acute bronchitis (466.0) -Acute upper respiratory infections (465.9) -Bronchitis (490) -Acute sinusitis (461.9)	-Azithromycin -Promethazine-codeine -Guaifenesin-codeine -Proair HFA -Levofloxacin	-X-ray chest -Chest PA & Lateral -Pulse oximetry -Serv prov during reg sched eve/wkend/hol hrs -Inhalation Rx for obstruction MDI/NEB

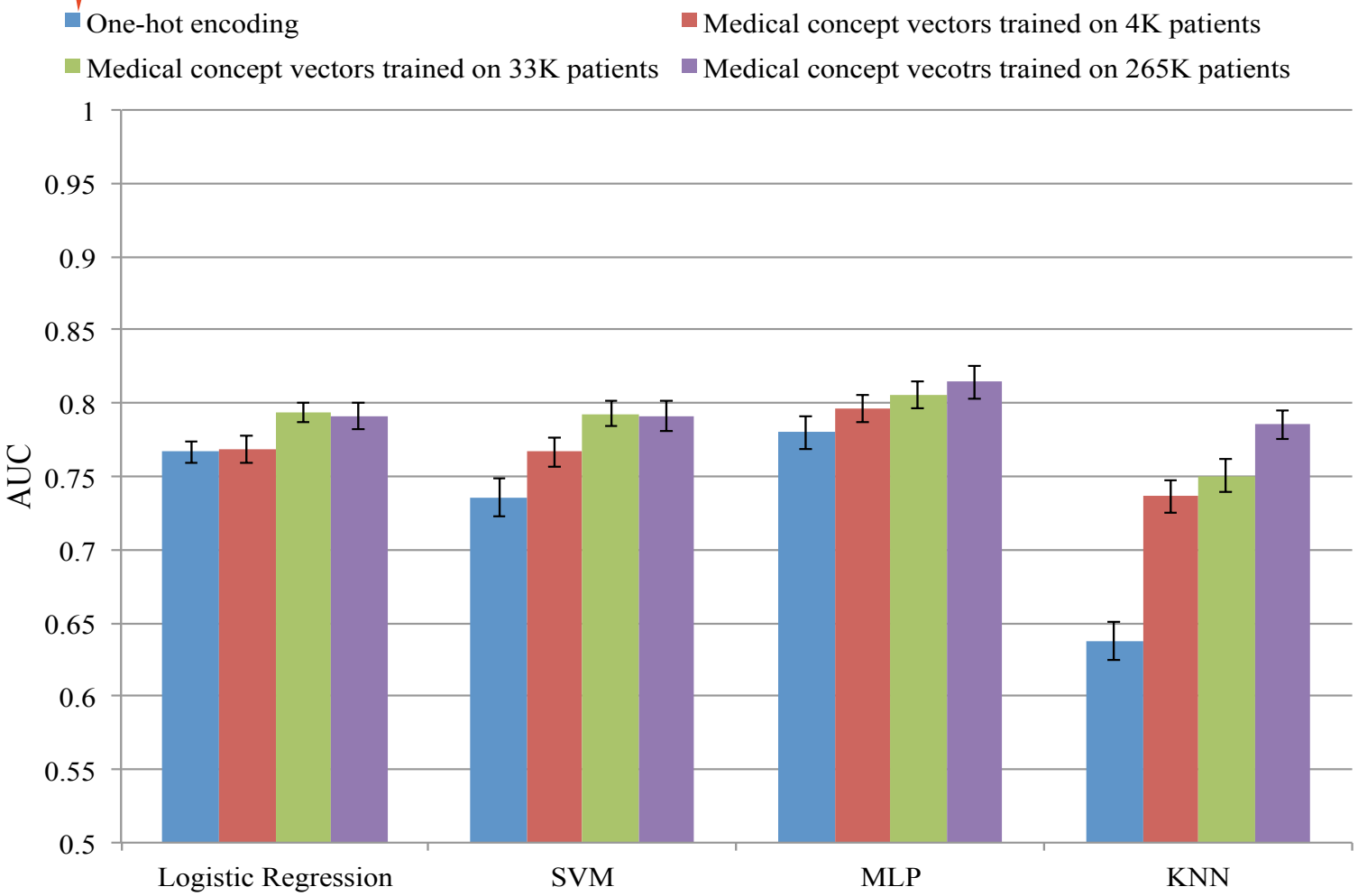
ILLINOIS

Algebraic operation on word2vec embeddings

	Diagnoses	Medications	Procedures
Loss of Weight (783.21) + Anxiety State (300.00)	-Depressive disorder (311) -Malaise & fatigue (780.79) -Insomnia (780.52) -Generalized anxiety disorder (300.02) -Esophageal reflux (530.81)	-Lorazepam -Zolpidem tartrate -Omeprazole -Alprazolam -Trazodone HCL	-Referral to GI -ECG & Interpretation -GI -EKG -Chest PA & Lateral

ILLINOIS

Predictive modeling using word2vec embedding



Word2vec leads to more accurate prediction performance

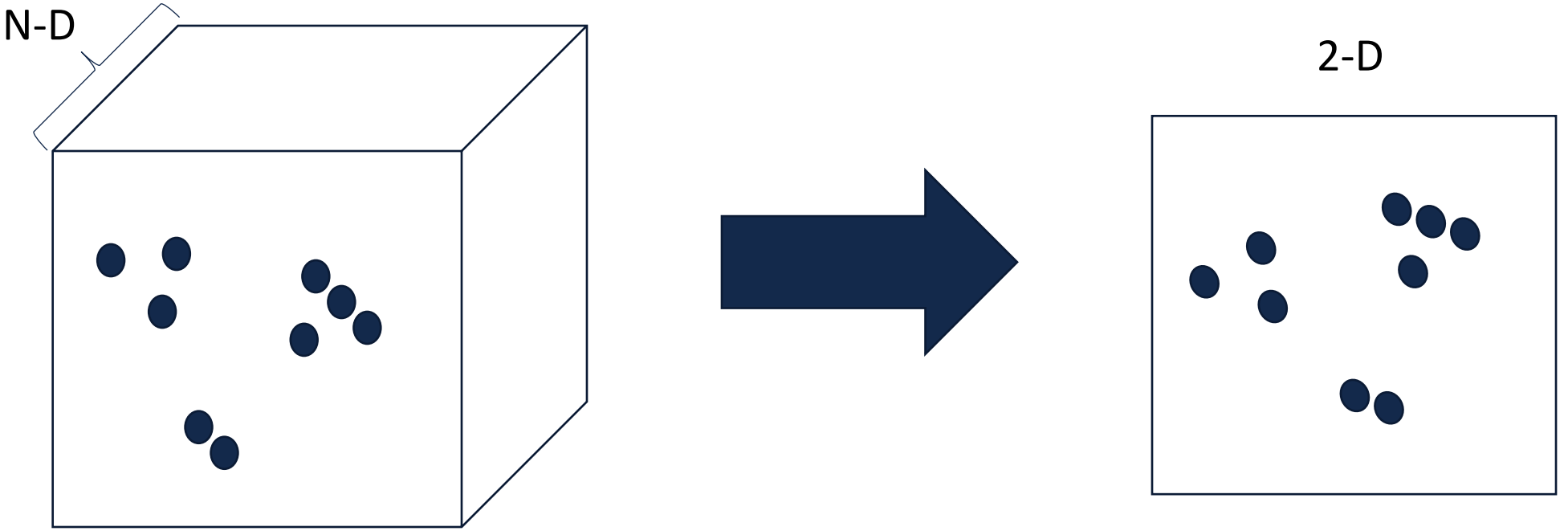
ILLINOIS

Visualizing high-dimensional data using t-SNE

General method

L.J.P. van der Maaten and G.E. Hinton. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research* 9(Nov):2579-2605, 2008.

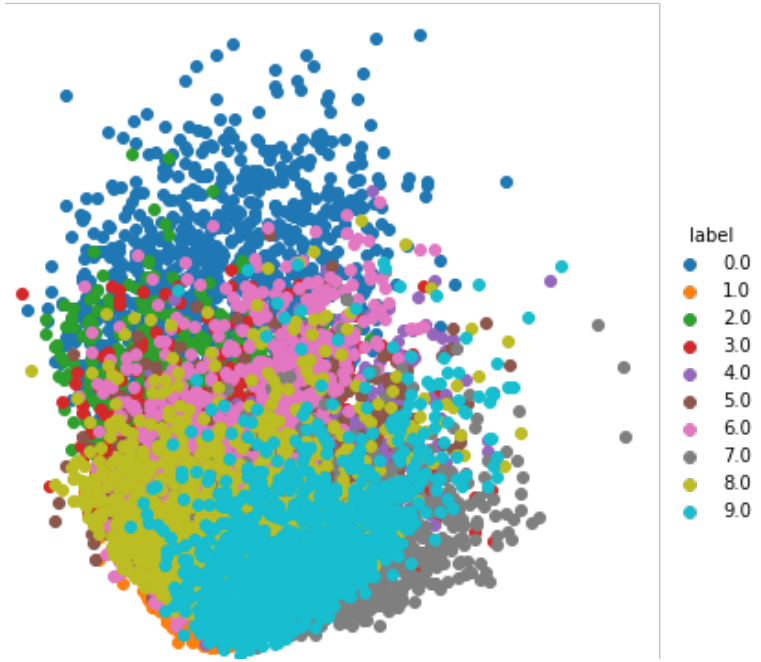
Motivation: How to visualize high-dimensional data?

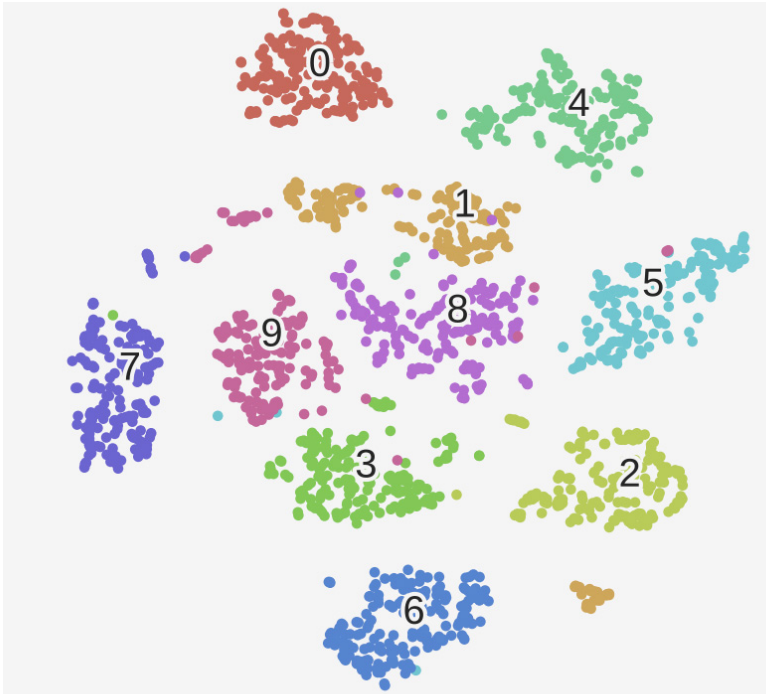


- Map high-dimensional data to low dimensional space while **preserving important properties**

ILLINOIS

What properties to preserve?

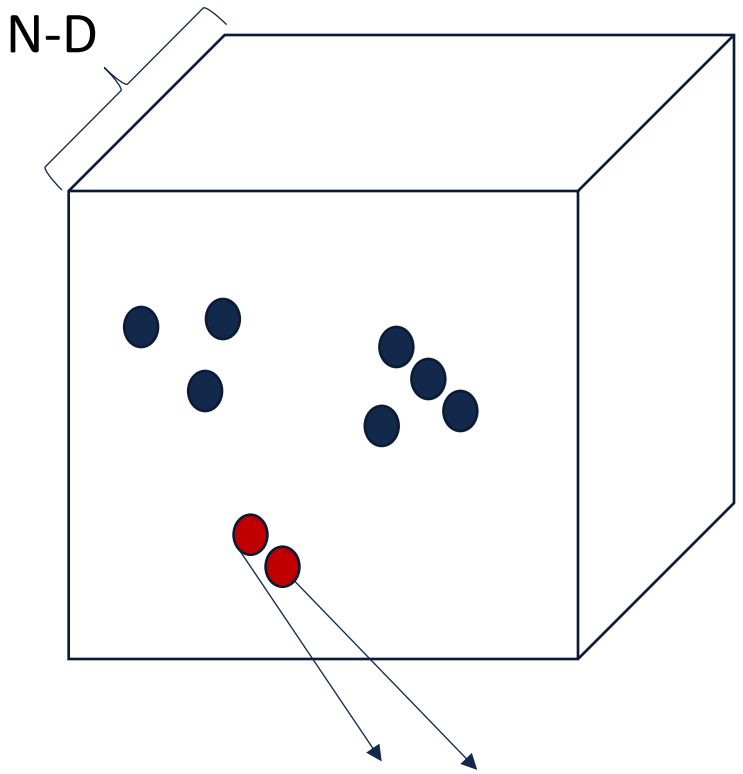




- Preserve pairwise distance ➔ Principal component analysis (PCA)
  - Faraway points will dominate the embedding
- Preserve local neighborhoods ➔ t-distributed Stochastic neighbor embedding (t-SNE)
  - Nearby points will dominate the embedding

ILLINOIS

t-SNE: input distribution



- Joint distribution with Gaussian kernel normalized by all pairs
- Global bandwidth σ is hard to set
- Alternative way is to use conditional distribution and set bandwidth σ_i for each point i
- And a simple hack to make p_{ij} symmetric

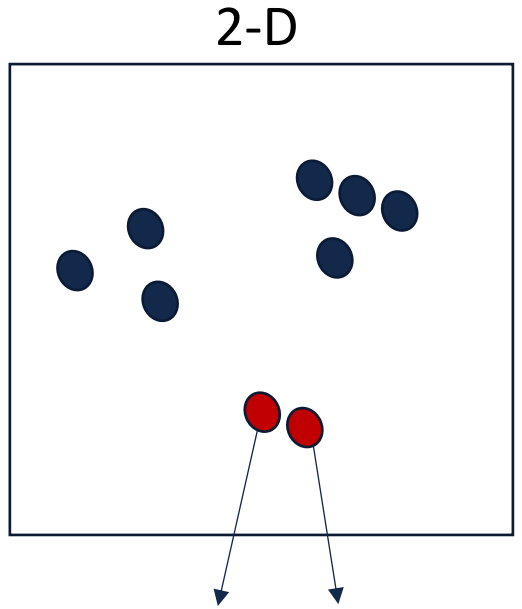
$$p_{ij} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)}{\sum_{k\neq l}\exp(-\|\mathbf{x}_k - \mathbf{x}_l\|^2/2\sigma^2)}$$

$$\begin{cases} p_{j|i} = \dfrac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma_i^2)}{\sum_{k\neq i}\exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2/2\sigma_i^2)} \\ p_{ij} = (p_{j|i} + p_{i|j})/2 \end{cases}$$

ILLINOIS

t-SNE: output distribution

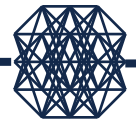
- The output distribution is student t-distribution with 1 degree of freedom
- It is a heavier tailed distribution than input distribution (Gaussian).
  - Benefit:
    - smaller distance from input can be mapped to larger distance in output
    - leads to better space utilization for visualization



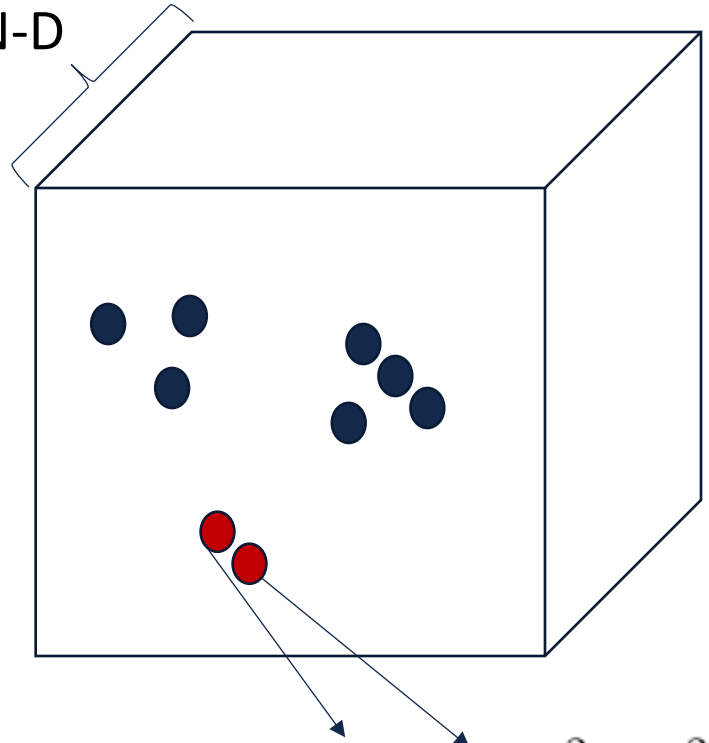
$$q_{ij} = \frac{(1 + \|\mathbf{y}_i - \mathbf{y}_j\|)^{-1}}{\sum_{k \neq l}(1 + \|\mathbf{y}_k - \mathbf{y}_l\|)^{-1}}$$

ILLINOIS

t-SNE: objective



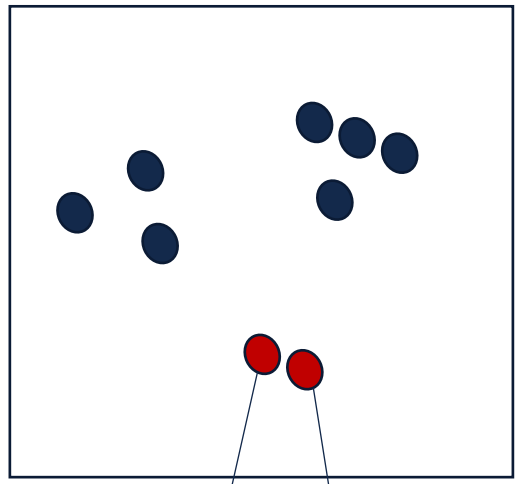
N-D



KL divergence is used

$$KL(P||Q) = \sum_i \sum_{j \neq i} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

2-D



$$p_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2 / 2\sigma_i^2)}$$

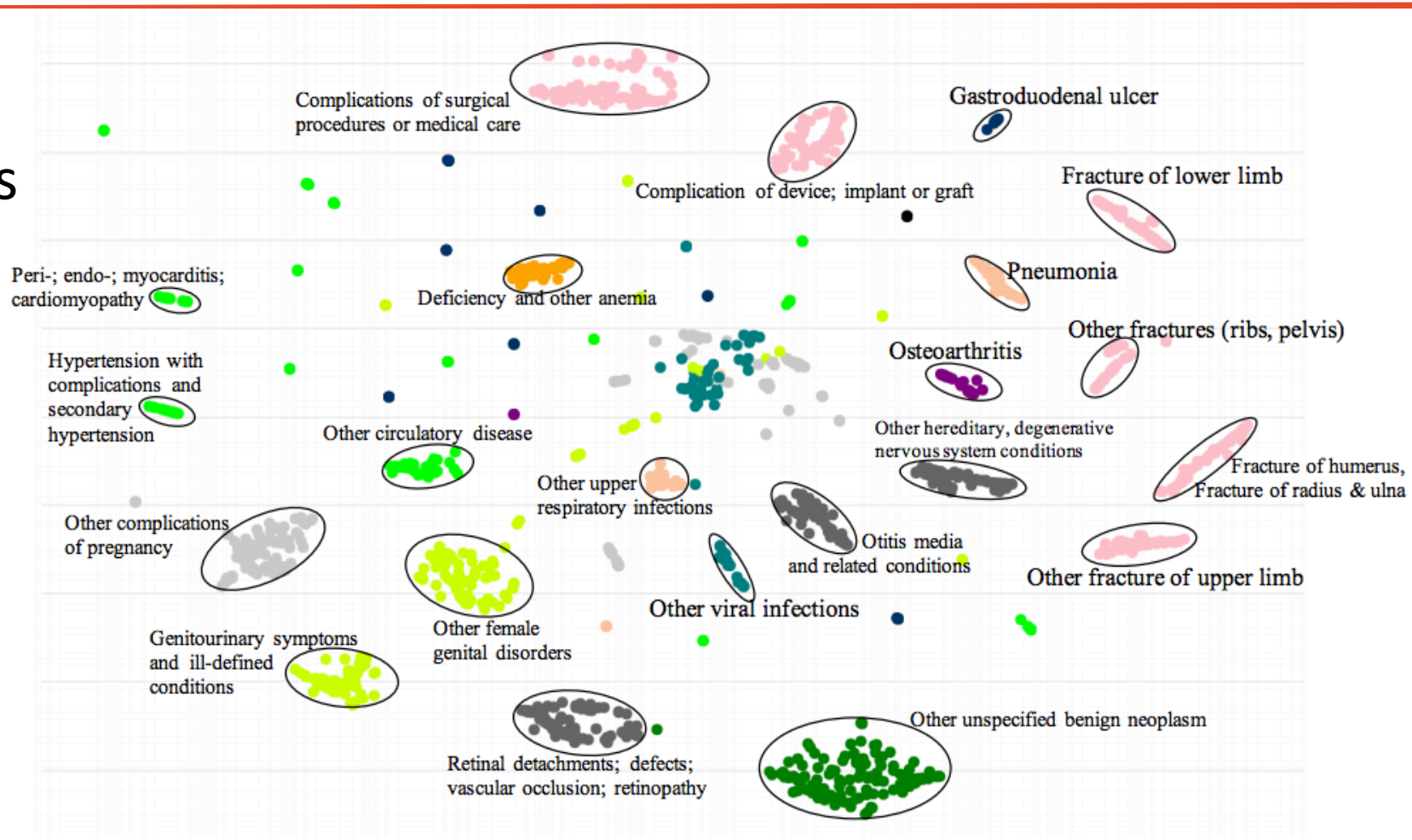
$$p_{ij} = (p_{j|i} + p_{i|j})/2$$

$$q_{ij} = \frac{(1 + \|\mathbf{y}_i - \mathbf{y}_j\|)^{-1}}{\sum_{k \neq l} (1 + \|\mathbf{y}_k - \mathbf{y}_l\|)^{-1}}$$

ILLINOIS

Result of t-SNE

Disease phenotype clusters are visualized

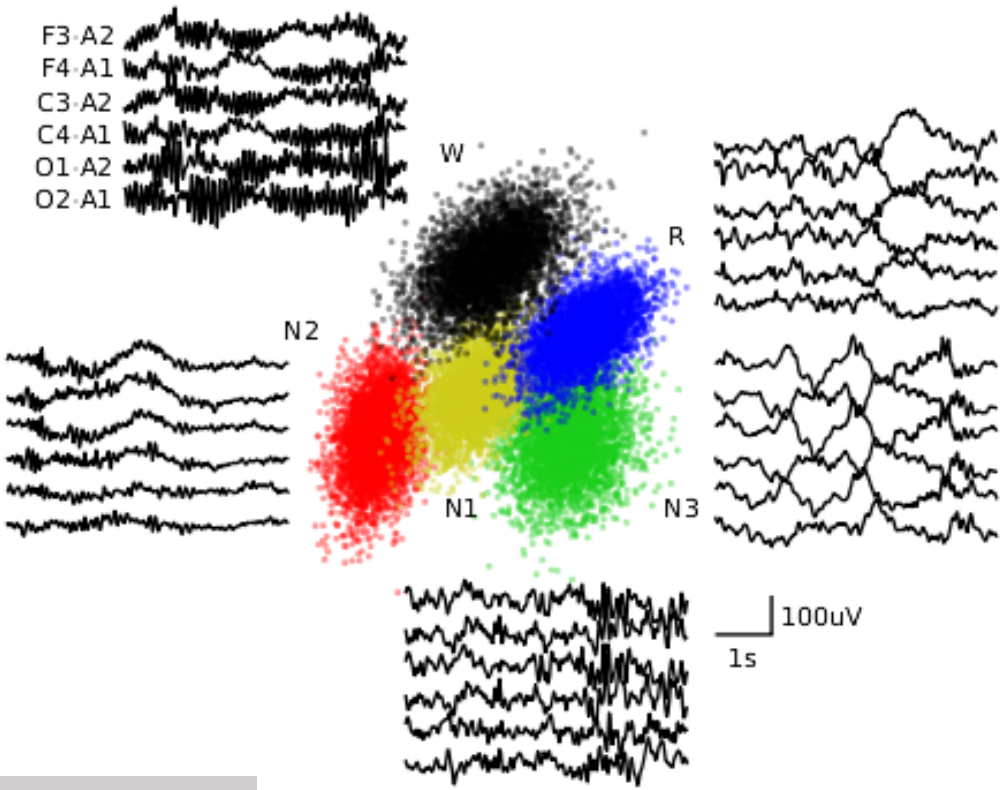


Choi, Edward, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, and Jimeng Sun. 2017. "GRAM: Graph-Based Attention Model for Healthcare Representation Learning." KDD

ILLINOIS

T-SNE plot of sleep stages

- Different sleep stages are visualized



Biswal, Siddharth, Haoqi Sun, Balaji Goparaju, M. Brandon Westover, Jimeng Sun, and Matt T. Bianchi. 2018. “Expert-Level Sleep Scoring with Deep Neural Networks.” *Journal of the American Medical Informatics Association: JAMIA*

ILLINOIS

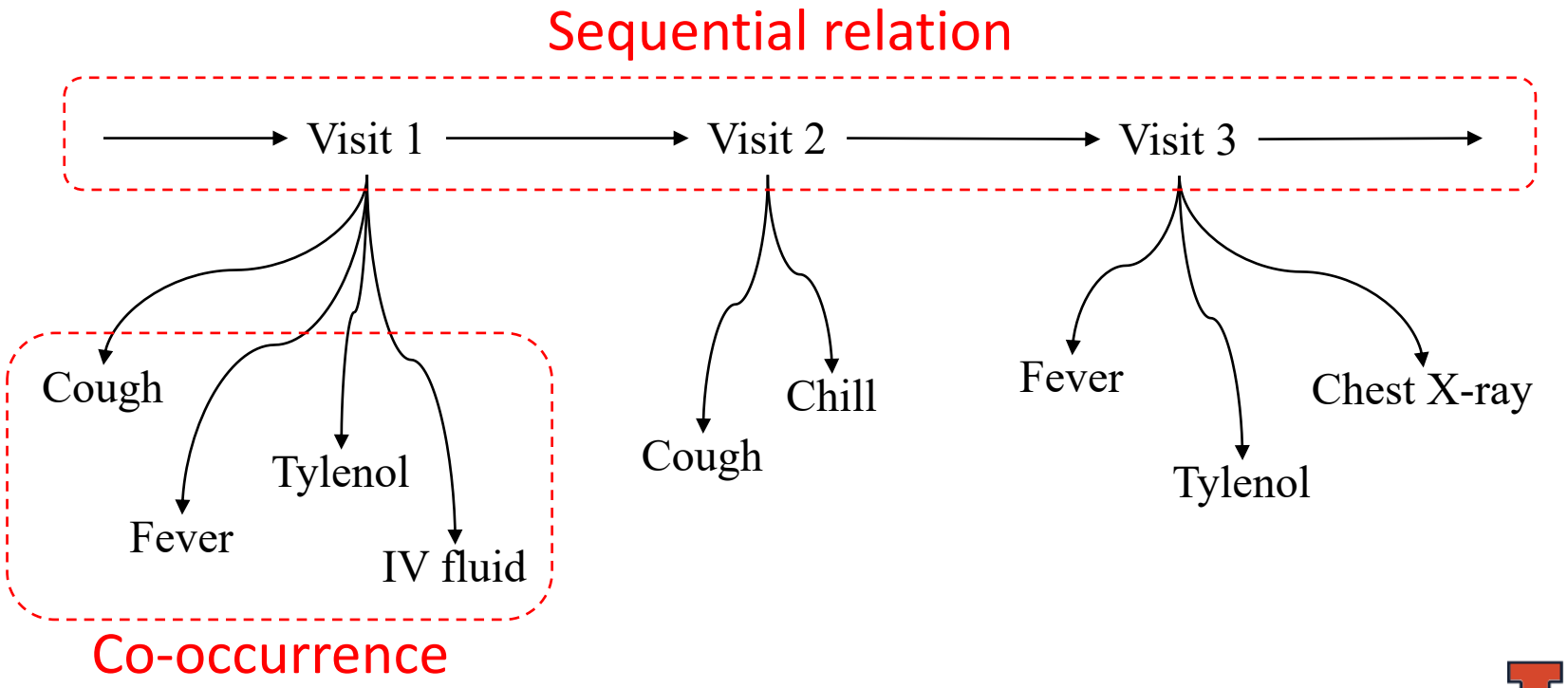
Med2Vec: Multi-layer Representation Learning for Medical Concepts

Edward Choi, Mohammad T. Bahadori, Elizabeth Searles,
Catherine Coffey, Jimeng Sun

KDD'16

Med2Vec: Background

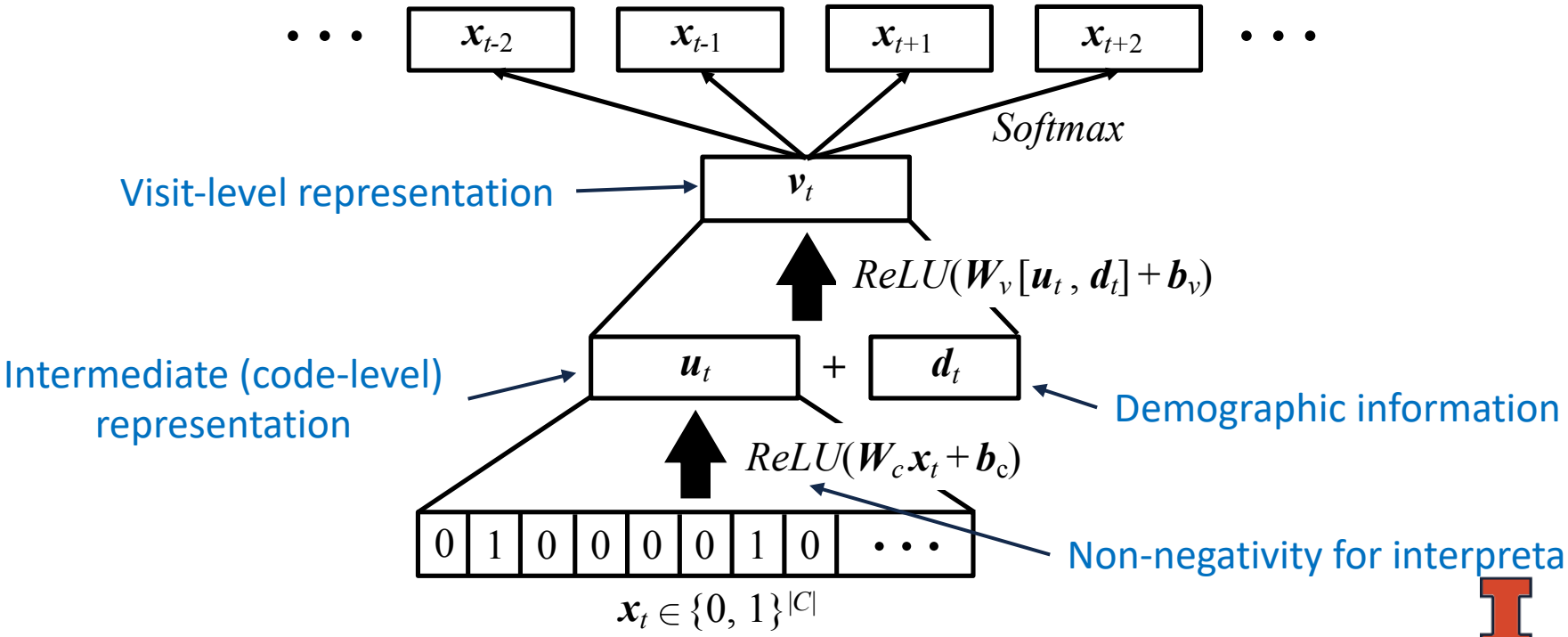
- Learn good representations of medical concepts
  - Diagnosis/medication/procedure codes
- Utilize 2-level structure of EHR



ILLINOIS

Med2Vec: Model

- Model architecture
  - Exploit two-layer structure of longitudinal EHR
    - Intra-visit codes provide co-occurrence information
    - Visit sequence provides sequential information



Med2Vec: Data

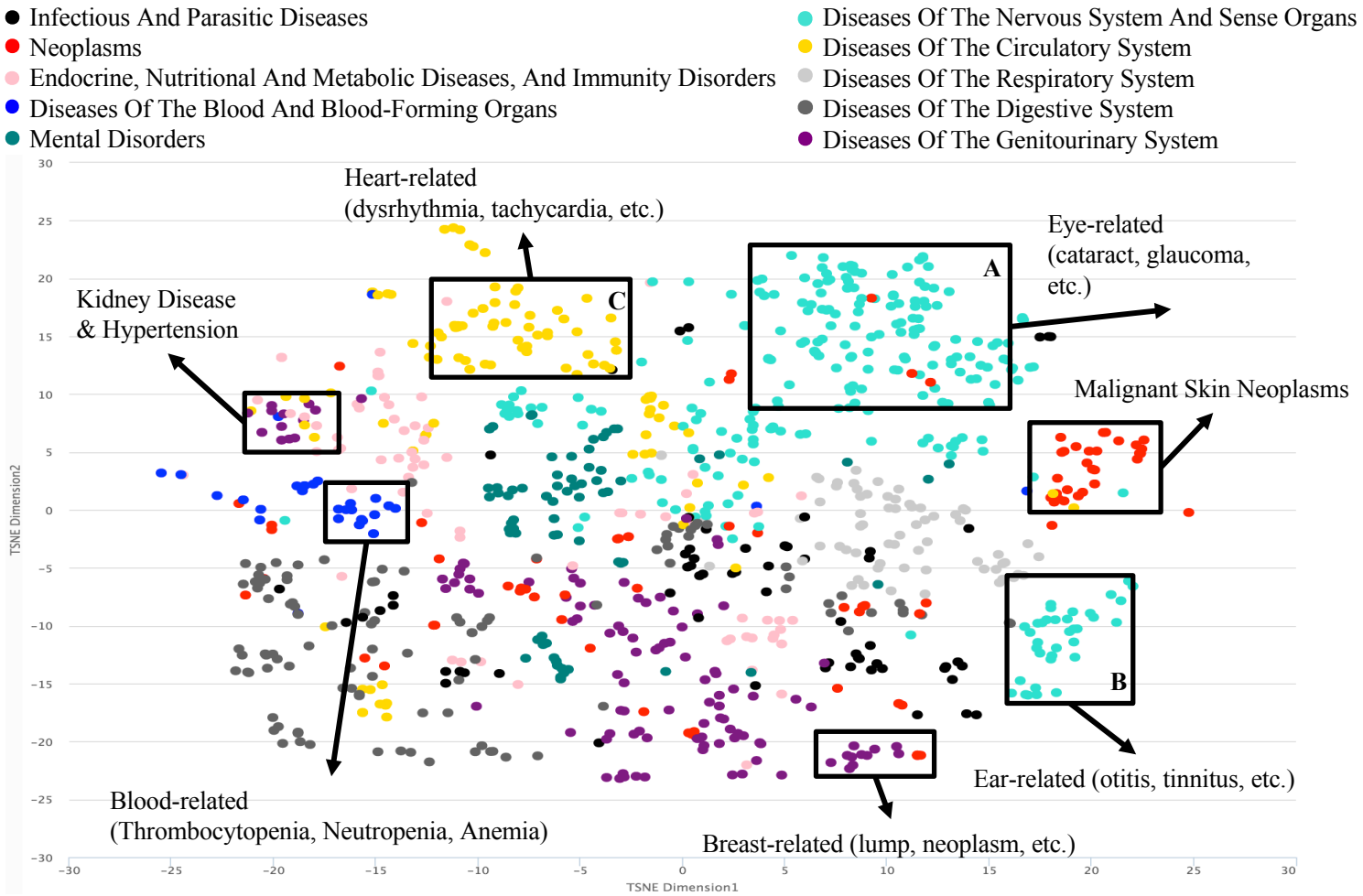
From Children's Healthcare of Atlanta (CHOA)

Dataset	CHOA
# of patients	550,339
# of visits	3,359,240
Avg. # of visits per patient	6.1
# of unique medical codes	28,840
- # of unique diagnosis codes	10,414
- # of unique medication codes	12,892
- # of unique procedure codes	5,534
Avg. # of codes per visit	7.88
Max # of codes per visit	440
(95%, 99%) percentile # of codes per visit	(22, 53)

ILLINOIS

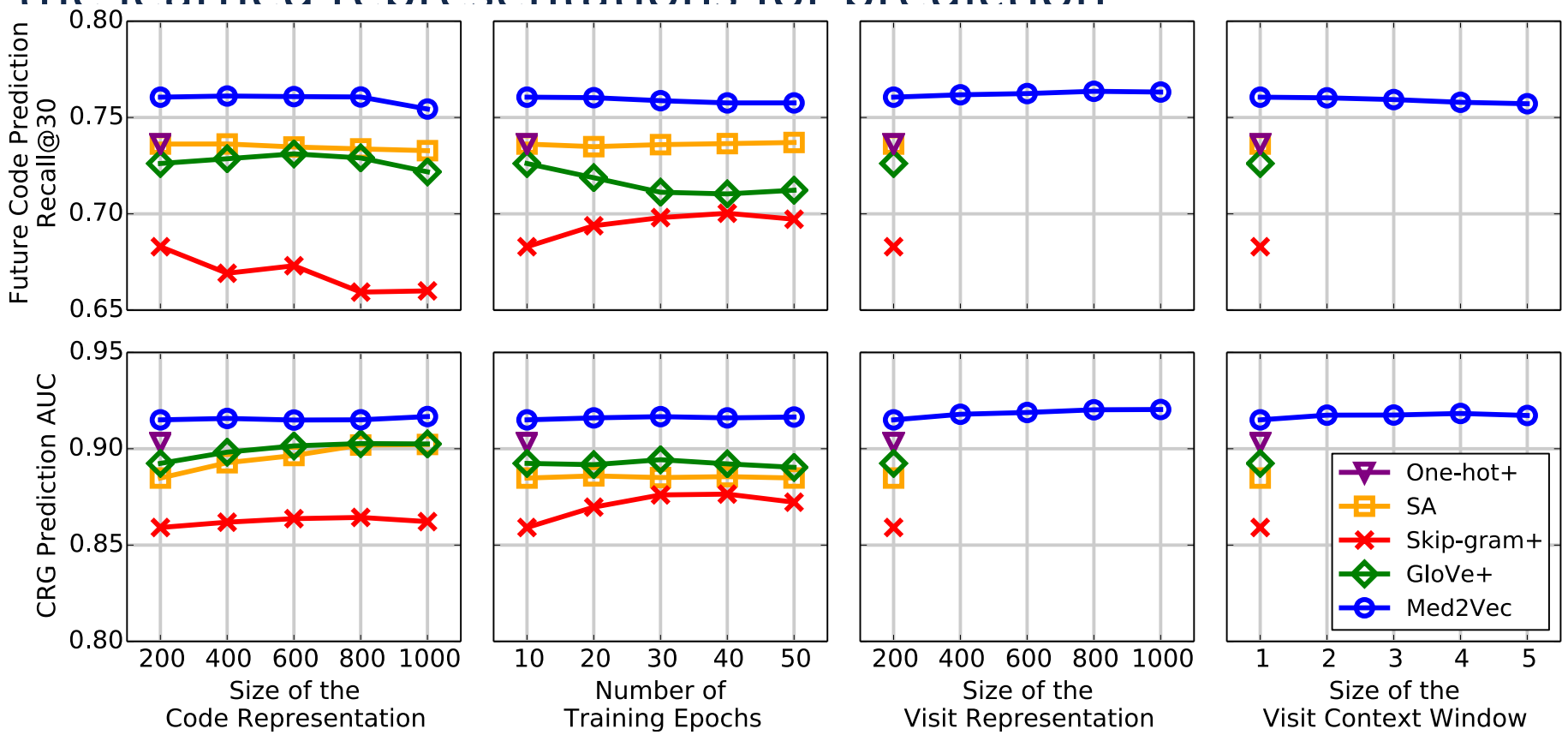
Med2Vec: Result

- Visualizing the learned representations



Med2Vec: Prediction

- Using the learned representations for prediction



ILLINOIS

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

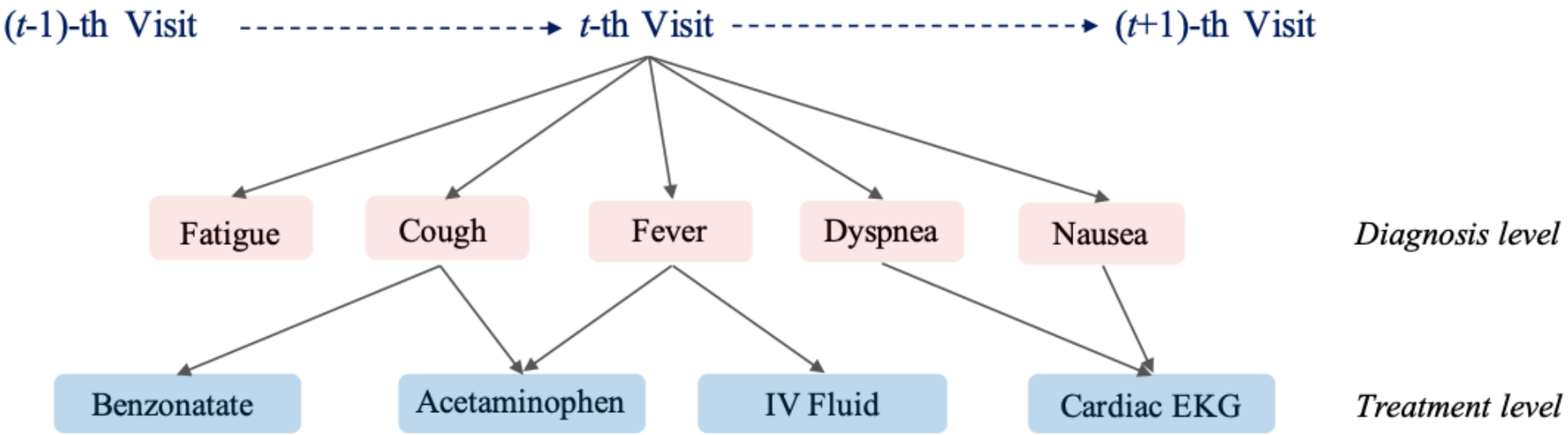
Edward Choi, Cao (Danica) Xiao, Walter Stewart, Jimeng Sun

NeurIPS'18

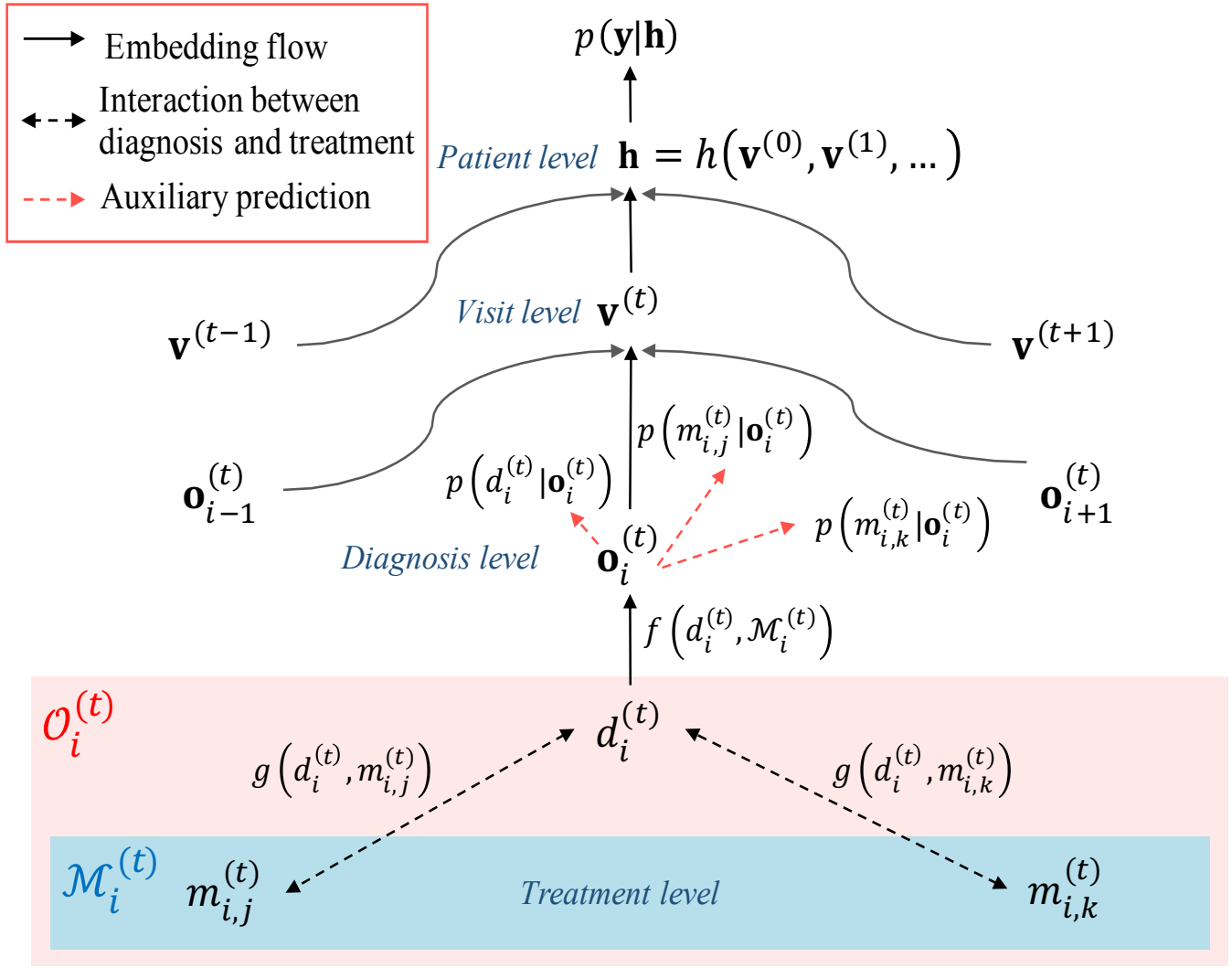
ILLINOIS

Leverage Structures within Electronic Health Records

- Temporal dependency across visits
- Hierarchical structure within a visit
  - Diagnosis treatment relations



MiME Architecture



4. Generate patient representation vector $\mathbf{h}$.

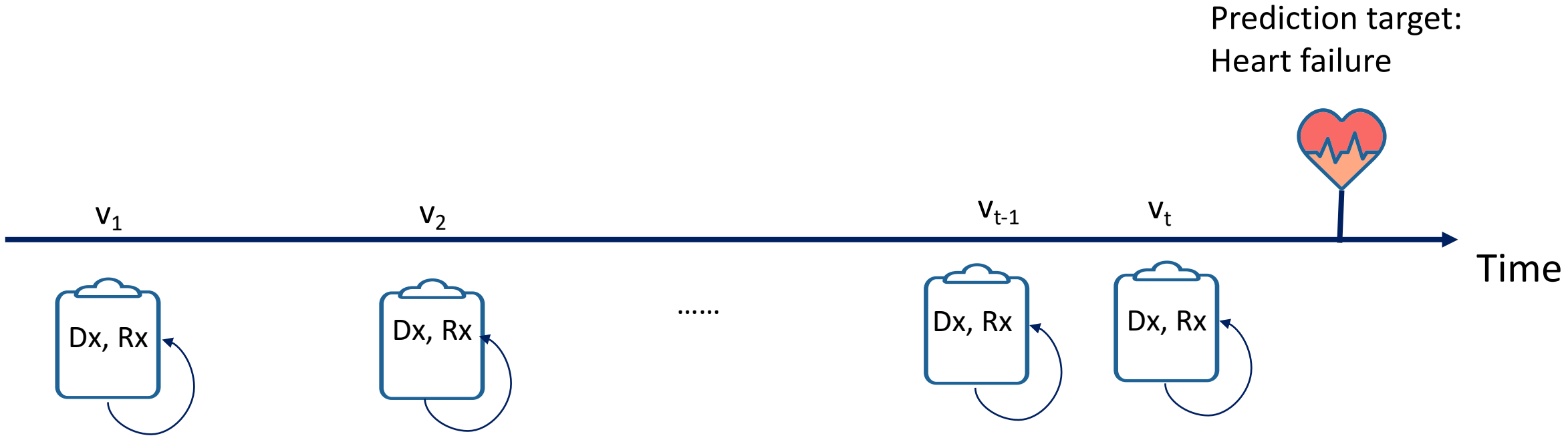
3. Generate representation vector $\mathbf{v}$ for t-th visit.

2. Generate representation vector $\mathbf{o}_i$ for i-th diagnosis.

1. Interaction between diagnosis d_i and its j-th treatment $m_{i,j}$

Auxiliary prediction task

- **Intuition**: to overcome limited labels by introducing ”free“ labels
- **Approach**: Additional supervised tasks on Dx and Rx prediction.



Experiments

# 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)

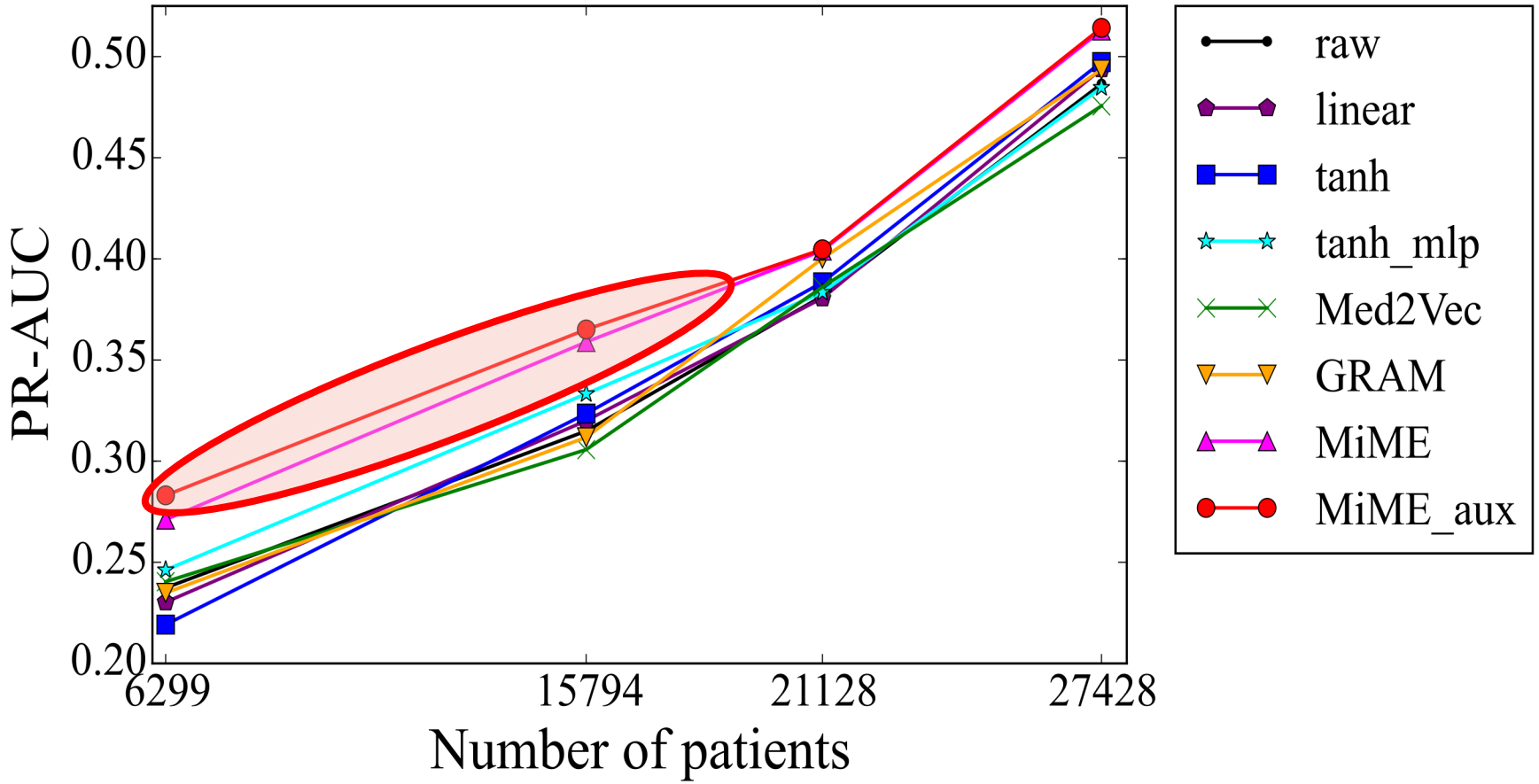
Source of data: Longitudinal EHR consisting of patients (age 50-85) from Sutter Health for heart failure studies.

Targets:

- Heart failure prediction

Results (Heart Failure Prediction)

- MiME outperforms all baselines specially for smaller cohorts
- Boost prediction power for rare conditions and new EHR systems



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

MiME learns a predictive patient representation

Capture relations within a visit

Introduce auxiliary task to enhance prediction

Demonstrated strong prediction performance

especially benefit smaller cohorts