

# Patient Representation using Autoencoder

Cinyoung Hur  
2018.7.7

I'm

Senior Researcher, Data Engineer  
working at [linewalks](#)

Co-organizer of



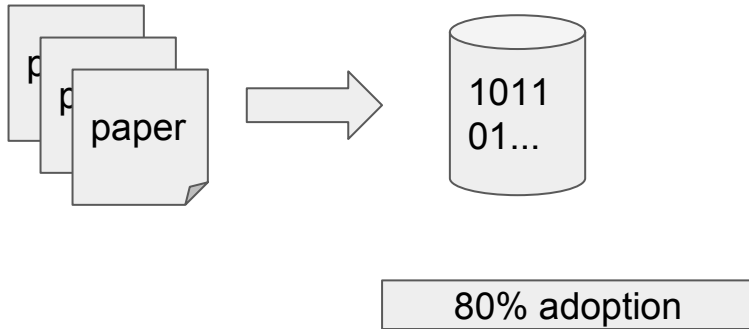
# Overview

- What is Electronic Health Record
- How to represent medical concepts and patient information
- Demo!

# Electronic Health Records

- Patient Registry
- Disease Registry
- Drug Registry

- Electronic Health Record
  - A collection of health data from all clinicians involved in a patient's care
  - more patient-centric, powerful and useful for diagnosis and treatment



# EHR data types

- Demographics
- Encounters
- Diagnosis
- Procedures
- Physical exams
- Sensor measurements
- Laboratory test results
- Prescribed or administered medications
- Clinical notes

# EHR data types

- Numerical quantities
  - body mass index, height, blood pressure
- Datetime
  - data of birth, time of admission
- Categorical value
  - ethnicity, codes from controlled vocabularies
- Natural language free-text
  - progress notes, discharge summaries
- Derived time series
  - vital sign signals during the course of the operation

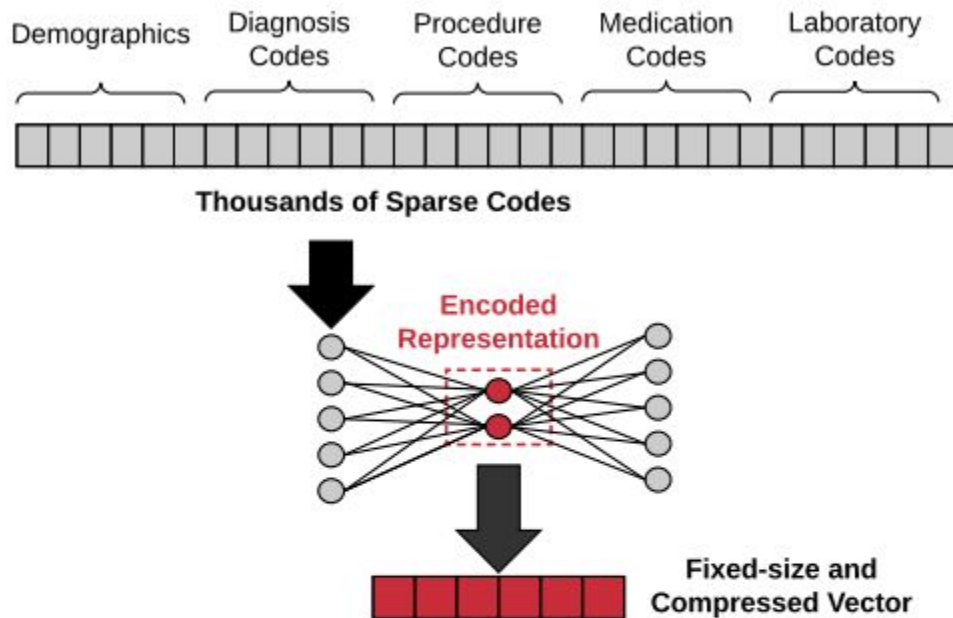
# Example Classification Schema

- Demographics
- Encounters
- **Diagnosis\***
- **Procedures\***
- Physical exams
- Sensor measurements
- **Laboratory test results\***
- **Prescribed or administered medications\***
- Clinical notes

Schema	Number of Codes	Examples
ICD-10 ( <i>Diagnosis</i> )	68,000	- J9600: Acute respiratory failure - I509: Heart failure - I5020: Systolic heart failure
CPT ( <i>Procedures</i> )	9,641	- 72146: MRI Thoracic Spine - 67810: Eyelid skin biopsy - 19301: Partial mastectomy
LOINC ( <i>Laboratory</i> )	80,868	- 4024-6: Salicylate, Serum - 56478-1: Ethanol, Blood - 3414-0: Buprenorphine Screen
RxNorm ( <i>Medications</i> )	116,075	- 161: Acetaminophen - 7052: Morphine - 1819: Buprenorphine

# EHR Representation Learning

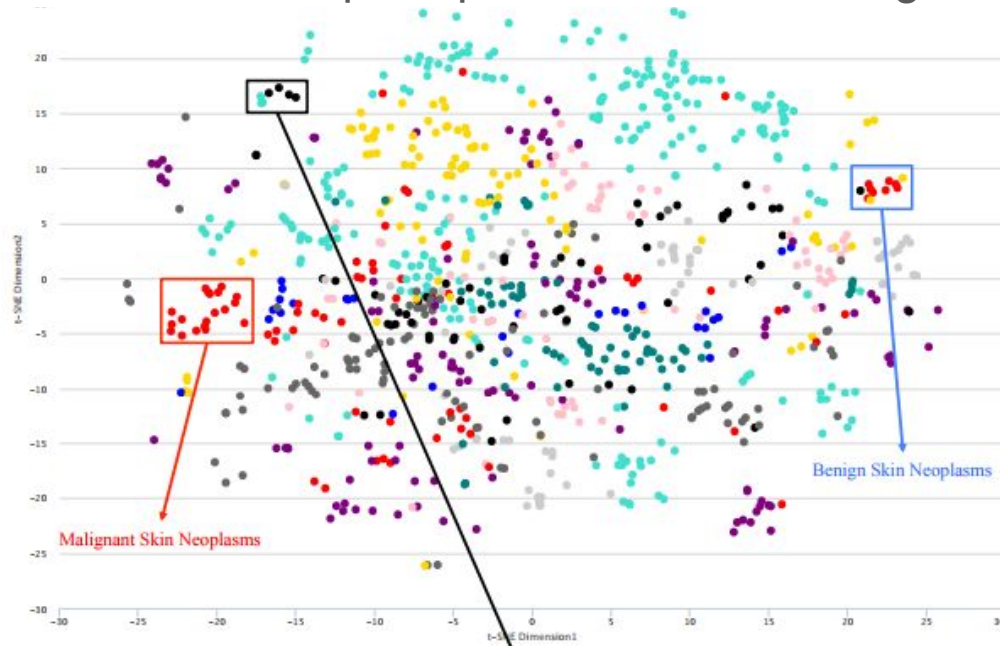
- Project discrete codes into vector space





# EHR Representation Learning : Concept Representation

- Evaluation of Medical Concept Representation Learning



Demo!

# MIMIC III dataset

- Medical Information Mart for Intensive Care III
- a large, freely-available database comprising de-identified health-related data
- over forty thousand patients who stayed in critical care units in the Beth Israel Deaconess Medical Center between 2001 and 2012.

## MIMIC Code Repository build passing DOI 10.5281/zenodo.821872 chat on glitter

This is a repository of code shared by the research community. The repository is intended to be a central hub for sharing, refining, and reusing code used for analysis of the [MIMIC critical care database](#). To find out more about MIMIC, please see: <https://mimic.physionet.org>

You can read more about the code repository in the following open access paper: [The MIMIC Code Repository: enabling reproducibility in critical care research](#).

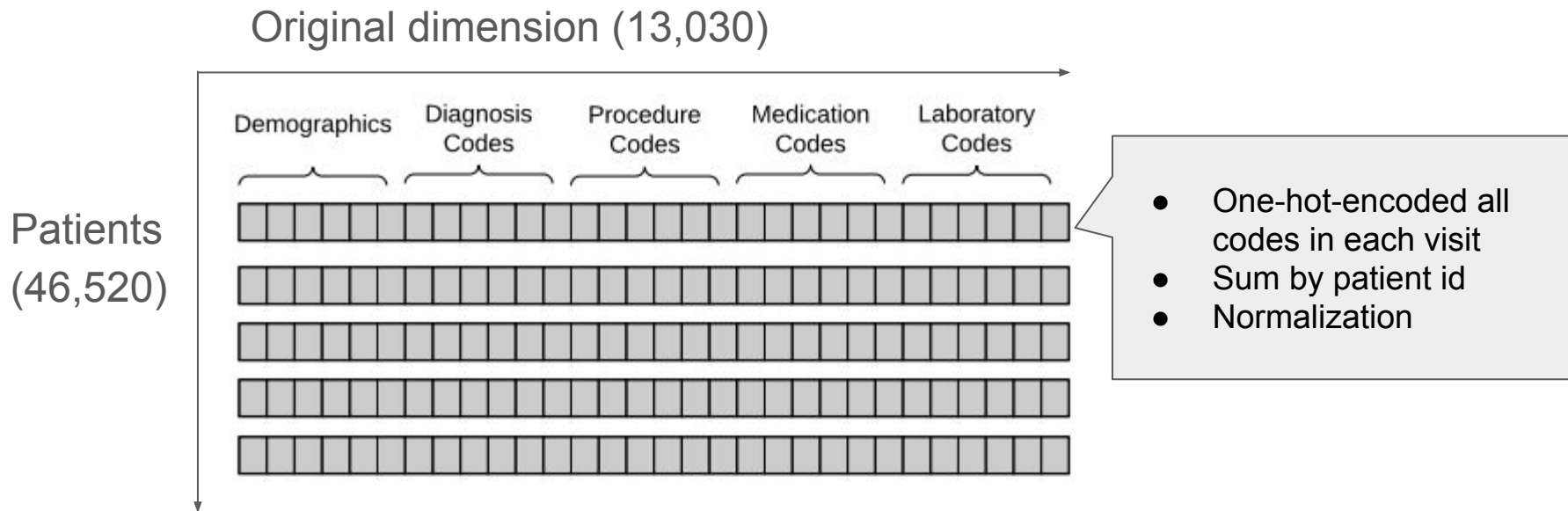
If you use code or concepts available in this repository, we would be grateful if you would cite the above paper as follows:

Johnson, Alistair EW, David J. Stone, Leo A. Celi, and Tom J. Pollard. "The MIMIC Code Repository: enabling reproducibility in critical care research." Journal of the American Medical Informatics Association (2017): ocx084.

You can also directly cite the repository using the above DOI from Zenodo.

<https://github.com/MIT-LCP/mimic-code>

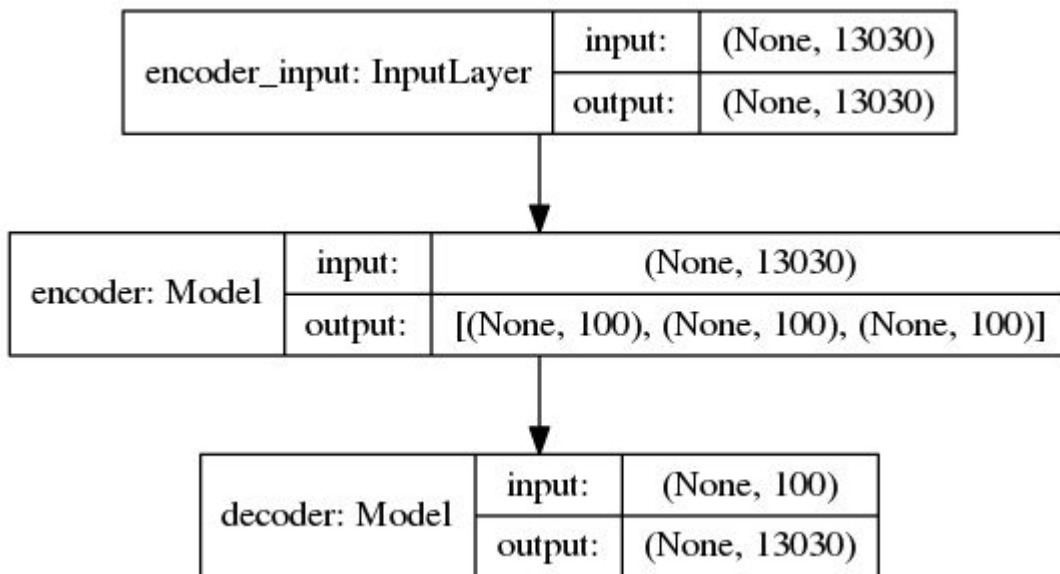
# MIMIC III Preprocessing



# Variational Autoencoder

- Keras

- original\_dim = 13030
- intermediate\_dim = 2048
- batch\_size = 100
- latent\_dim = 100
- epochs = 10



# We are looking for colleagues! Join us.



Linewalks is a healthcare big data startup that develops data services using machine learning and visualization techniques in large-scale data processing systems.

2016. Angel investing from KITE Entrepreneurship Foundation

2018. Series A investing from Kakao Investment (19 billion won)

- Positions
  - Frontend Developer
  - Backend Developer
  - Machine Learning Developer
- For more information, visit <http://bit.ly/lw-tech-recruit>

Thanks!