香港浸會大學
HONG KONG BAPTIST UNIVERSITY

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

Joint Learning of Phenotypes and Diagnosis-Medication Correspondence via Hidden Interaction Tensor Factorization

Kejing Yin[1], William K. Cheung[1], Yang Liu[1], Benjamin C. M. Fung[2], Jonathan Poon[3]

[1] Department of Computer Science, Hong Kong Baptist University
[2] School of Information Studies, McGill University
[3] Hong Kong Hospital Authority

Background

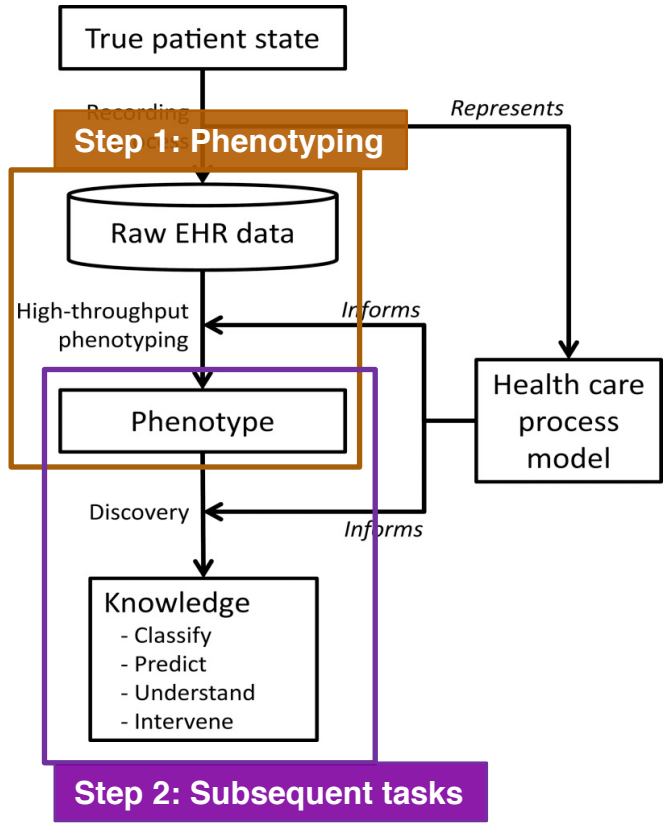
DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

- **Phenotyping from Electronic Health Records (EHR)**

  The process of mapping raw EHR data into clinically meaningful and relevant features (phenotypes).

- **Using raw EHR data is challenging**
  - High degree of missing and inaccuracy.
  - Highly complex nature of healthcare and possible bias of clinicians.
  - Raw EHR data cannot directly reflect the patients' health states.

- **Two-step Approach of using EHR data:**
  1. Transform the raw EHR data to clinically relevant features via phenotyping.
  2. Use the resulting phenotypes as features for the subsequent tasks.



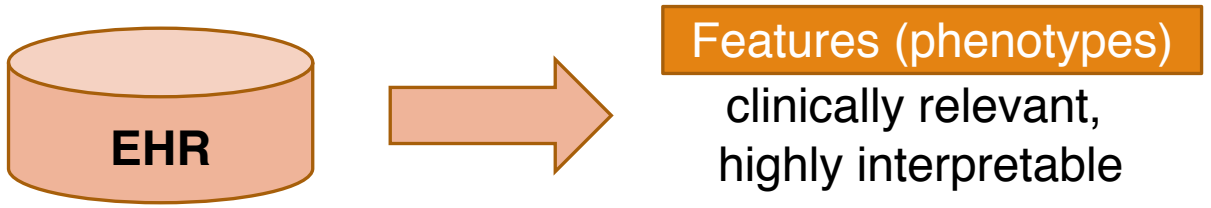
[1] Kirby, Jacqueline C., et al. "PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability." *Journal of the American Medical Informatics Association* 23.6 (2016): 1046-1052.
[2] Ho, Joyce C., et al. "Limestone: High-throughput candidate phenotype generation via tensor factorization." *Journal of biomedical informatics* 52 (2014): 199-211.
[3] Yang, Kai, et al. "TaGiTeD: Predictive Task Guided Tensor Decomposition for Representation Learning from Electronic Health Records." *AAAI*. 2017.

Background

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

- **Phenotyping from Electronic Health Records (EHR)**

  The process of mapping raw EHR data into clinically meaningful and relevant features (phenotypes).

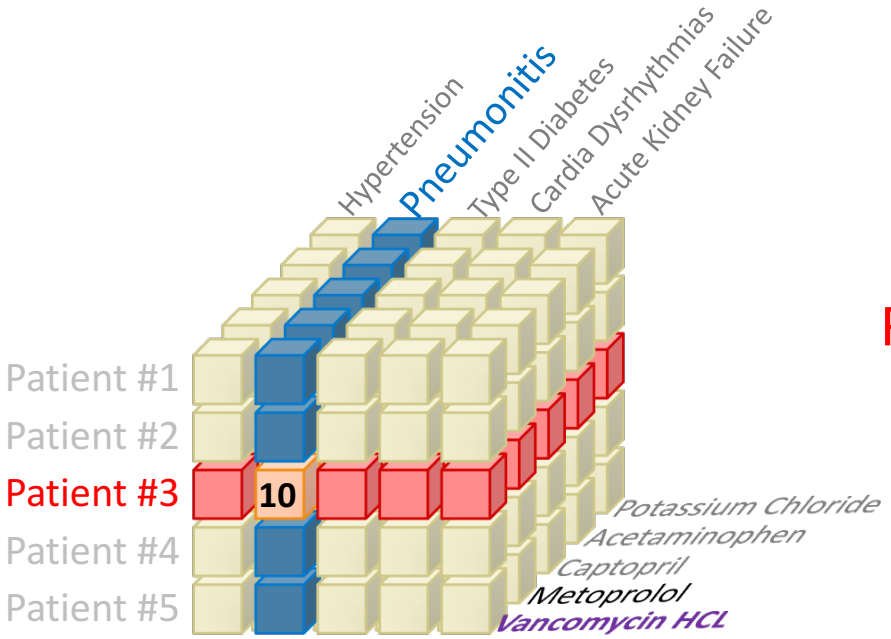


- **Computational Phenotyping**
  - Phenotyping without intensive human supervision. (unsupervised problem)

Tensor Factorization as a Data-Driven Method

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

- Several studies have explored using tensor factorization for computational phenotyping[1-6].
- Tensor is a higher-order extension of matrix, where the high-order interactions among EHR data can be naturally represented, e.g.



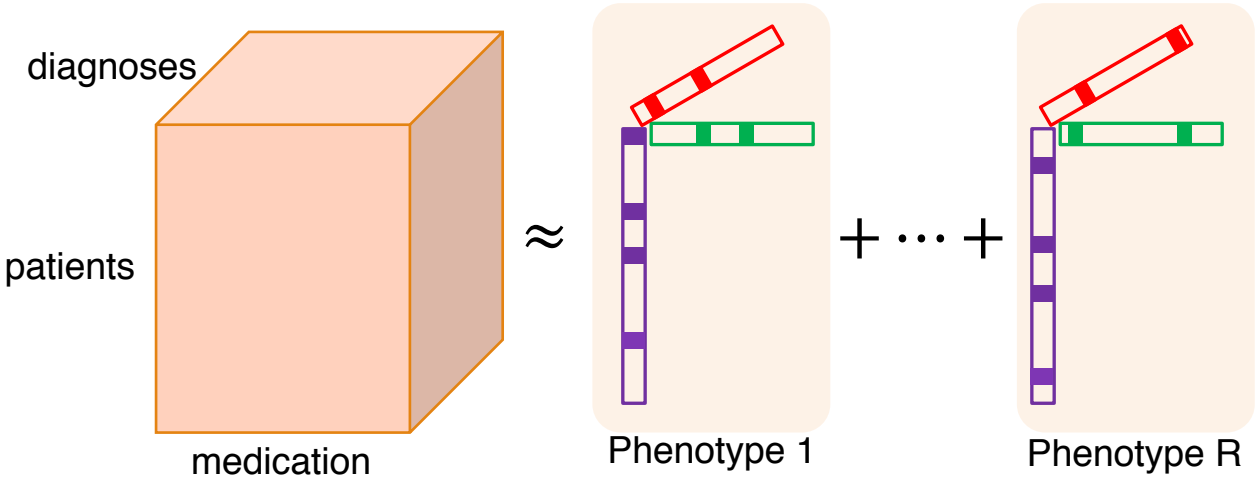
Patient #3 is prescribed with Vancomycin HCL for ten times in response to Pneumonitis.

[1] Ho, Joyce C., et al. "Limestone: High-throughput candidate phenotype generation via tensor factorization." *Journal of biomedical informatics* 52 (2014): 199-211.
[2] Ho, Joyce C., Joydeep Ghosh, and Jimeng Sun. "Marble: high-throughput phenotyping from electronic health records via sparse nonnegative tensor factorization." *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 2014.
[3] Wang, Yichen, et al. "Rubik: Knowledge guided tensor factorization and completion for health data analytics." *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 2015.
[4] Kim, Yejin, et al. "Discriminative and distinct phenotyping by constrained tensor factorization." *Scientific reports* 7.1 (2017): 1114.
[5] Yang, Kai, et al. "TaGiTeD: Predictive Task Guided Tensor Decomposition for Representation Learning from Electronic Health Records." *AAAI.* 2017.
[6] Henderson, Jette, et al. "Granite: Diversified, Sparse Tensor Factorization for Electronic Health Record-Based Phenotyping." *2017 IEEE International Conference on Healthcare Informatics (ICHI),* 2017.

Tensor Factorization as a Data-Driven Method

DEPARTMENT OF COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

- Non-negative CP factorization for computational phenotyping:



Approximation with sum of R rank-one tensors:

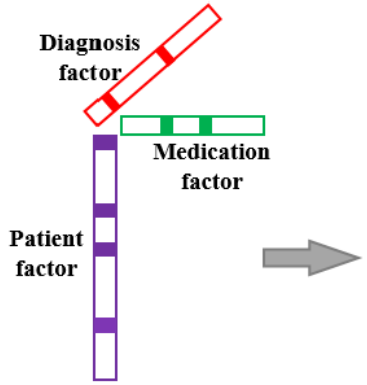
$$\mathcal{X} \approx \hat{\mathcal{X}} = \sum_{r=1}^{R} a_r \circ b_r \circ c_r$$

Minimize the reconstruction error: $\min \ \text{Error}(\mathcal{X}, \hat{\mathcal{X}})$

Nature of tensor factorization: a low-rank model.

Global interaction patterns are captured by the rank-one tensors.

- Phenotype extraction from rank-one tensor:



Diagnosis factor

Diagnosis entry	value
Cardiac dysrhythmias	**0.7**
Diabetes mellitus	0
Heart failure	**0.3**
Asthma	0
Pneumonitis	0

Medication factor

Medication entry	value
Insulin	0
Metoprolol	**0.55**
Furosemide	**0.45**
Albuterol	0
Diltiazem	0

Desired Properties of phenotypes for interpretability:

- More sparsity (less element in one phenotype)
- Less similarity (less similar between each other)

[1] Kolda, T. G., & Bader, B. W. (2008). Tensor Decompositions and Applications. *SIAM Review*, 51(3)
[2] Chi, Eric C., and Tamara G. Kolda. On tensors, sparsity, and nonnegative factorizations. *SIAM Journal on Matrix Analysis and Applications* 33.4 (2012): 1272-1299.

Related Works

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

- Existing models differ from each other in their data distribution assumptions and constraints.

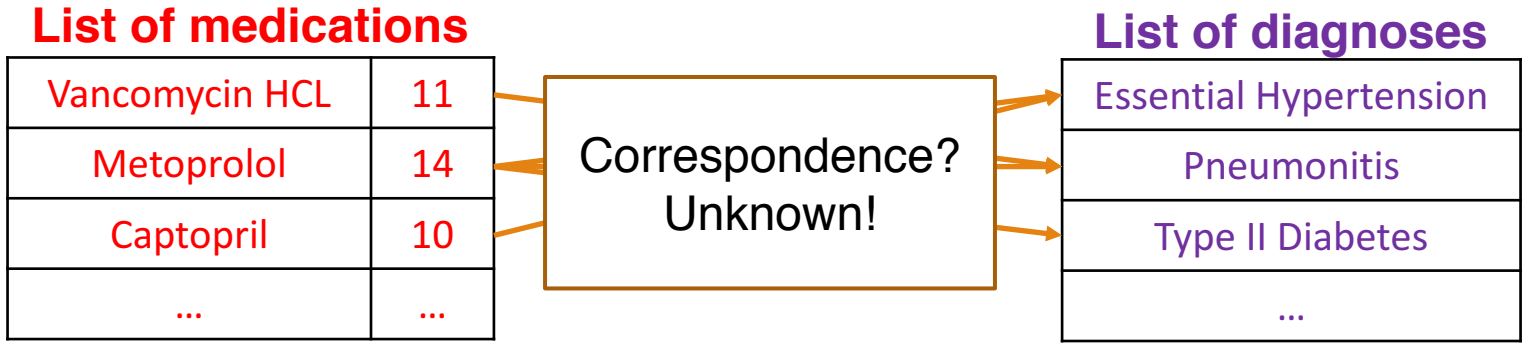
Model	Data distribution assumption and reconstruction error	Additional constraints
Limestone (J. Biomed. Inform., 2014)	Poisson, KL divergence	No
Marble (KDD, 2014)	Poisson, KL divergence	Adding bias term (capturing global information)
Rubik (KDD, 2015)	Gaussian, Frobenius norm	1. Guidance constraint (consistent with medical knowledge) 2. Pairwise constraint (making phenotypes more distinct)
Kim et al. (Scientific Report, 2017)	Gaussian, Frobenius norm	1. Prediction task (for improving discriminative power) 2. Clustering structure (making phenotypes distinct)
Kim et al. (KDD, 2017)	Gaussian, Frobenius norm	Distributed learning from multiple sites (without sharing patient level data)
TaGiTed (AAAI, 2017)	Poisson, KL divergence	Prediction task (improving predictive performance)
Granite (ICHC, 2017)	Poisson, KL divergence	1. Angular regularization (making phenotypes distinct) 2. Sparsity regularization

Research Challenges

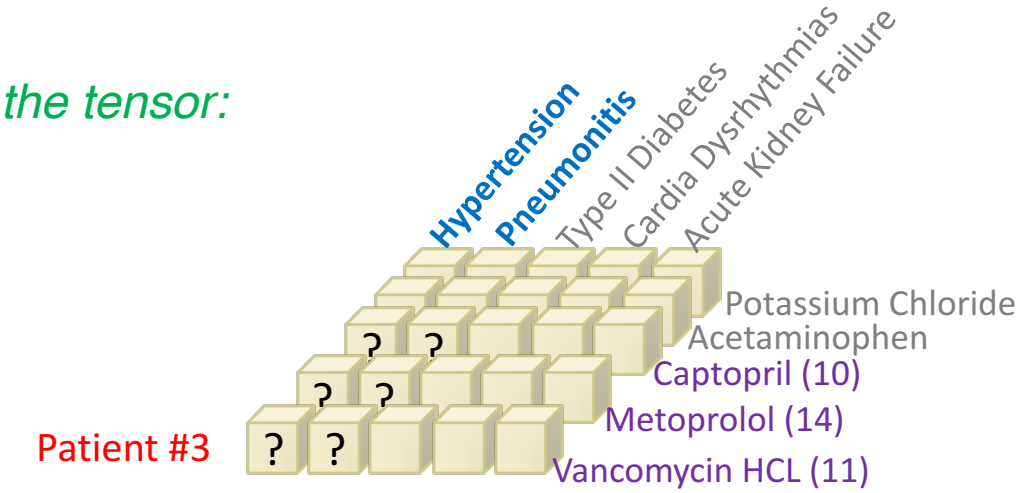
DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

- **Interaction information are often missing in the records.**

For example: a patient's records during a hospital admission:



Construction of the tensor:



How to fill in the entries?

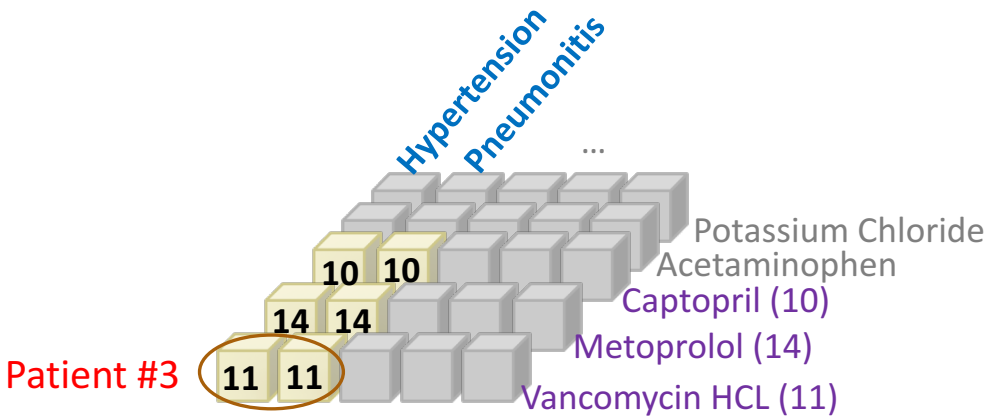
How to factorize the tensor when we do not observe it?

Research Challenges

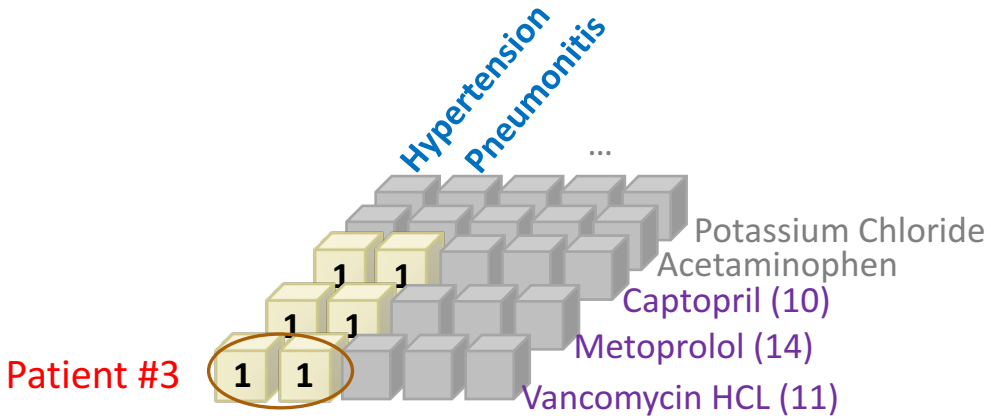
DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

- **Interaction information are often missing in the records.**

Existing models adopt the "equal-correspondence" strategy:



1. Count of co-occurrence



2. Binary Tensor

Underlying assumption: one medication corresponds to all co-occurring diagnoses equally.

The assumption could be unrealistic, e.g. vancomycin HCL is used for Pneumonitis, but typically not for hypertension in clinical practice.

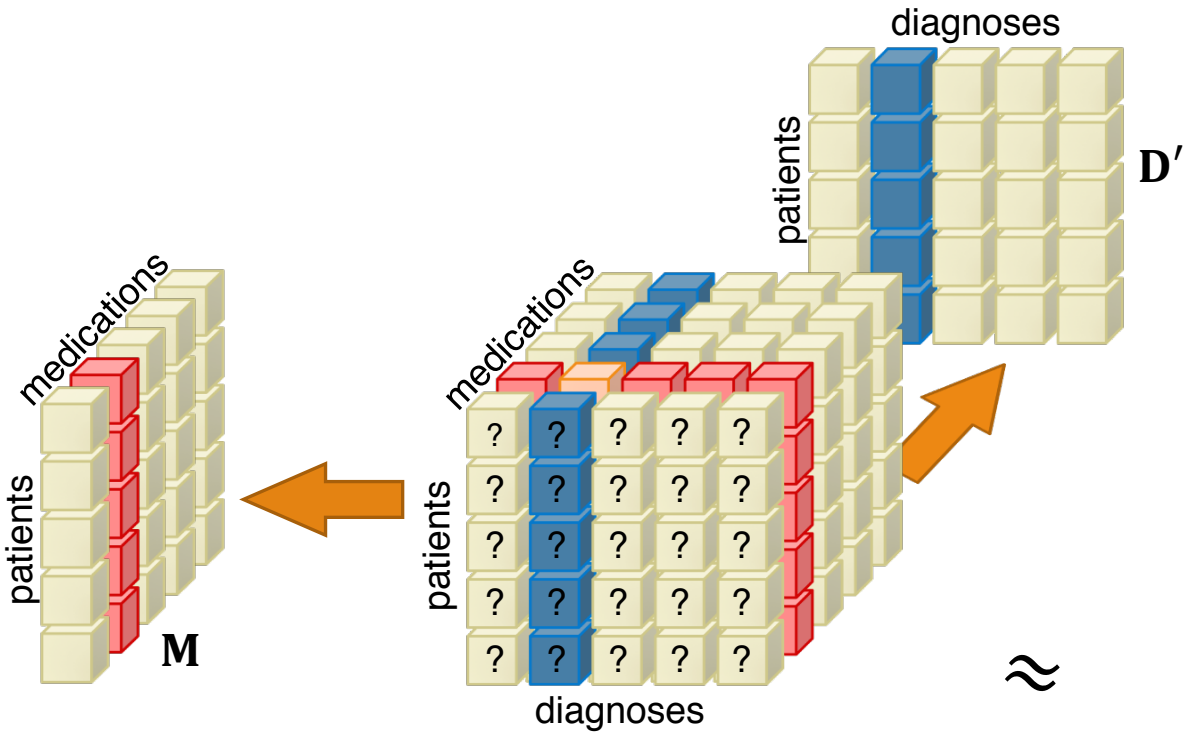
Recall: the objective of CP factorization is to **recover the interactions as much as possible**.

Inevitable error may be caused

Hidden Interaction Tensor Factorization

DEPARTMENT OF COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

Main Idea

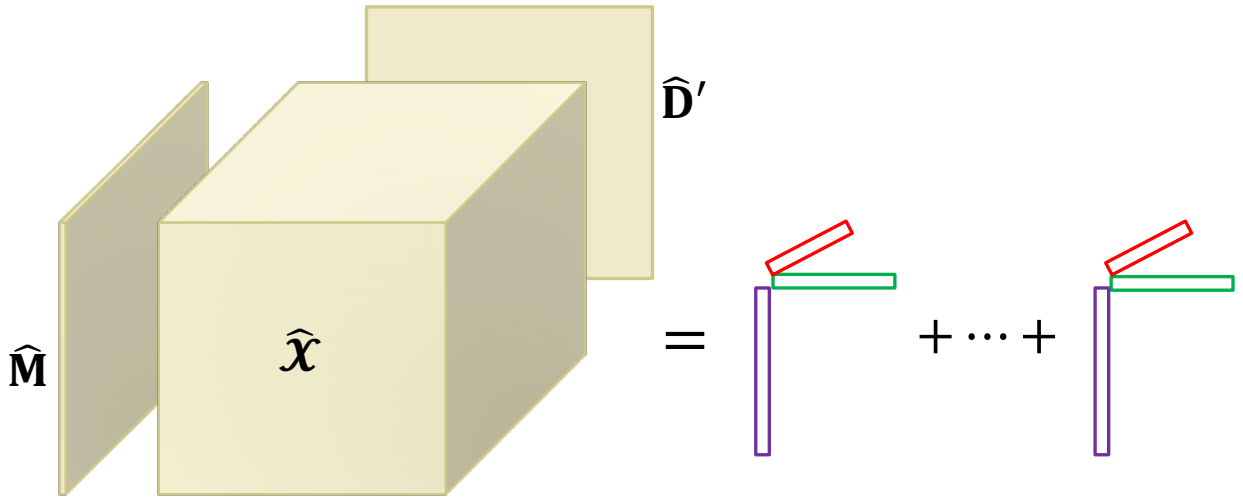


Interaction tensor $\mathcal{X}$: NOT observed

Instead of observing the interactions directly, we **observe the accumulation of the hidden interaction tensor**.

Approximate $\mathcal{X}$ by non-negative CP factorization

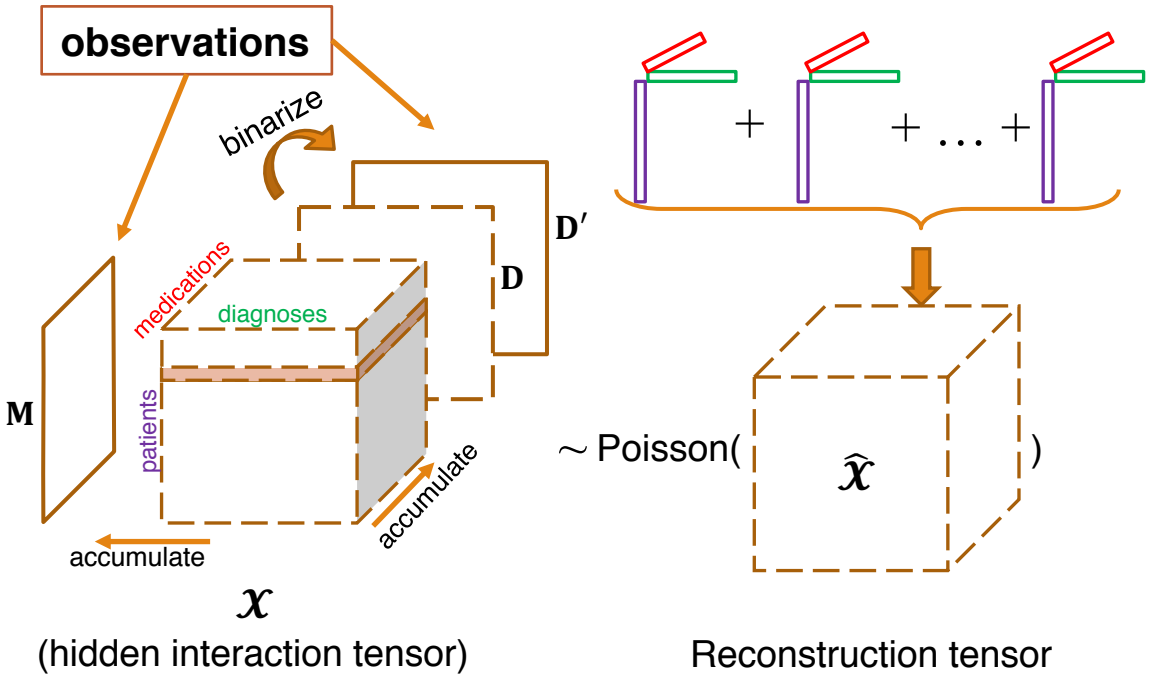
Instead of minimizing the error of reconstructing $\mathcal{X}$, we minimize the error of reconstructing $\mathbf{M}$ and $\mathbf{D}$.



Hidden Interaction Tensor Factorization

DEPARTMENT OF COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

Formulation



(hidden interaction tensor) Reconstruction tensor

Framework of **HITF**
(Hidden Interaction Tensor Factorization)

Poisson distribution for counting data: $x_{ijk} \sim \text{Poisson}(\hat{x}_{ijk})$

Sum of independent Poisson's yields also a Poisson:

$$m_{ik} = \sum_{j=1}^{N_d} x_{ijk} \sim \text{Poisson}(\sum_{j=1}^{N_d} \hat{x}_{ijk}) \qquad \mathbf{M} \sim \text{Poisson}(\mathbf{U}^{(1)} \text{diag}(\mathbf{1}^T \mathbf{U}^{(2)}) \mathbf{U}^{(3)^T})$$

For diagnoses, we observe only a binarization:

$$\Pr(d'_{ij} = 1) = 1 - \exp\left(-\sum_{r=1}^{R} \mathbf{u}_{ir}^{(1)} \left(\sum_{k=1}^{N_m} \mathbf{u}_{kr}^{(3)}\right) \mathbf{u}_{jr}^{(2)}\right) \quad \mathbf{D}' \sim \text{Ber}\left(1 - \exp\left(-\mathbf{U}^{(1)} \text{diag}(\mathbf{1}^T \mathbf{U}^{(3)}) \mathbf{U}^{(2)^T}\right)\right)$$

Log-Likelihood:

$$\begin{aligned}\mathcal{L} &= \mathcal{L}(\mathbf{M}) + \mathcal{L}(\mathbf{D}') = \sum_{i,k} \log\left(p\left(m_{ik} | \mathbf{U}^{(n)}\right)\right) + \sum_{i,j} \log\left(p\left(d'_{ij} | \mathbf{U}^{(n)}\right)\right) \\ &= \mathbf{1}^T(-\mathbf{U}^{(1)} \text{diag}(\mathbf{1}^T \mathbf{U}^{(2)}) \mathbf{U}^{(3)^T} + \mathbf{M} * \log(\mathbf{U}^{(1)} \text{diag}(\mathbf{1}^T \mathbf{U}^{(2)}) \mathbf{U}^{(3)^T}))\mathbf{1} \\ &+ \mathbf{1}^T(\mathbf{D}' * \log(\exp(\mathbf{U}^{(1)} \text{diag}(\mathbf{1}^T \mathbf{U}^{(3)}) \mathbf{U}^{(2)^T}) - \mathbf{E}) - \mathbf{U}^{(1)} \text{diag}(\mathbf{1}^T \mathbf{U}^{(3)}) \mathbf{U}^{(2)^T})\mathbf{1}.\end{aligned}$$

Infer the factor matrices by maximum likelihood estimation.

Learning Algorithms

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

Optimization Problem:

$$\begin{aligned} \underset{\mathbf{U}^{(1)},\mathbf{U}^{(2)},\mathbf{U}^{(3)}}{\arg\min} \quad & f(\mathbf{U}^{(1)},\mathbf{U}^{(2)},\mathbf{U}^{(3)}) \equiv -\mathcal{L}(\mathbf{M}) - \mathcal{L}(\mathbf{D}') \\ \text{subject to} \quad & \mathbf{U}^{(n)} \geq \mathbf{0}, n = 1,2,3. \end{aligned} \tag{13}$$

For each iteration,
fix all but one factor matrix, and update by solving the sub-problem.

E.g., for the patient dimension: $\mathbf{U}^{(1)}_{k+1} = \underset{\mathbf{X}\geq\mathbf{0}}{\arg\min} f(\mathbf{X}, \mathbf{U}^{(2)}_k, \mathbf{U}^{(3)}_k)$,

Apply block coordinate descent optimization:

Algorithm 1: Block Coordinate Descent Optimization Framework for HITF Model

Input : *patient-by-medication* matrix: $\mathbf{M}$,
patient-by-diagnoses matrix: $\mathbf{D}'$
Output: CP factor matrices: $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$

```
initialize U^(n) (n = 1, 2, 3) randomly;
repeat
    for n = 1 : 3 do
        repeat
            update U^(n) with other variables fixed using
              projected line search in Algorithm 2;
        until subproblem converges;
    end
until all subproblems converge;
```

Apply projected line search satisfying the Armijo condition:

Algorithm 2: Projected Line Search for Solving Sub-problems with Armijo Condition

Input : Variable $\mathbf{X}_k$, search direction $\mathbf{S}_k$,
sufficient descent σ and descent step ρ.
Output: Updated variable $\mathbf{X}_{k+1}$

```
t ← 0 ;
while not f(P_+[X_k + ρ^t S_k]) − f(X_k) ≤
  σ((P_+[X_k + ρ^t S_k] − X_k) · ∇f(X_k)) do
    t ← t + 1 ;
end
update variable: X_{k+1} ← P_+[X_k + ρ^t S_k];
```

Take negative gradient as the search direction: $\mathbf{S}_k = -\nabla f(\mathbf{X}_k)$

Experiments and Results

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

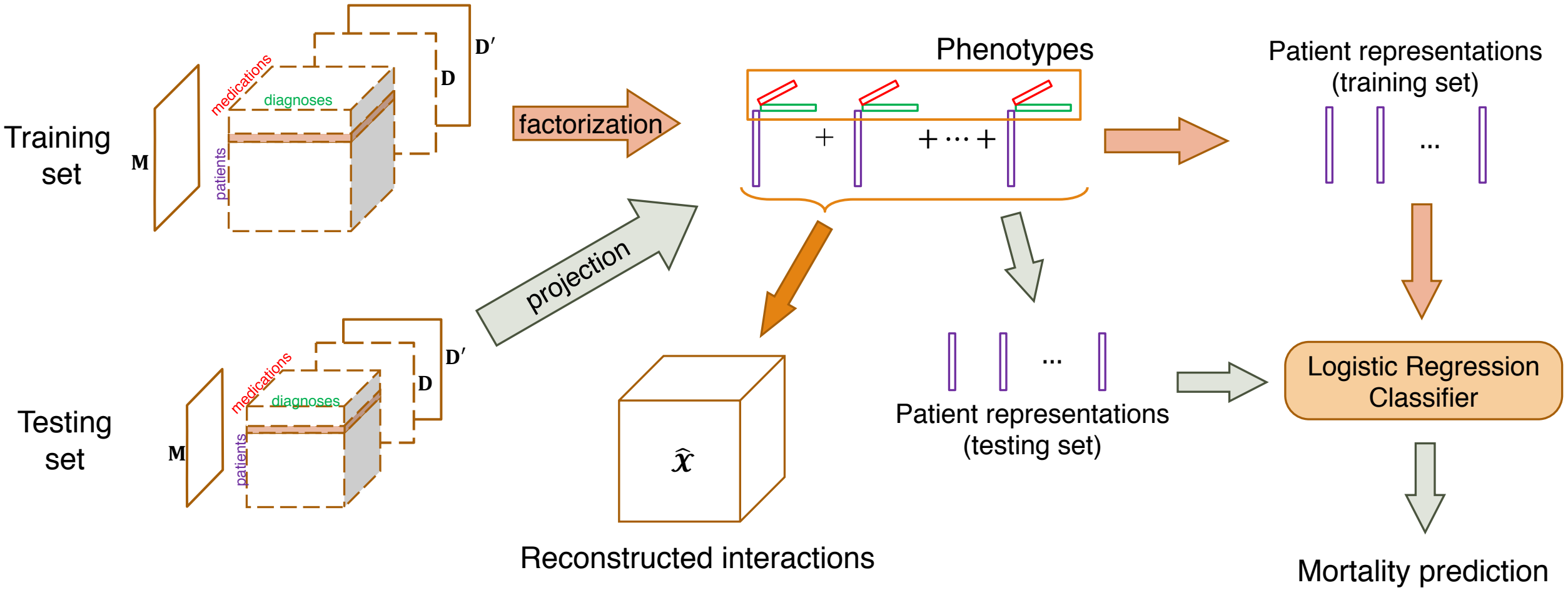
- **Data Set**:
  - MIMIC-III: open-source, large-scale, de-identified, ICU related.
  - Patients have 11 diagnoses per visit on average.
  - Contains many medications not for treating specific diseases (e.g. pain relievers).
- **Data Preprocessing**:
  - Exact a subset containing 7,652 adult patients with 50% died in hospital.
  - Use the first admission of each patient.
  - Exclude the base type drugs (e.g. D5W).
  - Use medications appeared in at least 5% of the patients (128 distinct ones).
  - Group ICD-9 code by the first three digits, and use those appeared in at least 1% of the patients (184 distinct ones.)
- **Baselines**:
  - Rubik: one of the state-of-the-art NTF computational phenotyping model.
  - CP-APR: a widely used Poisson NTF model.
  - SiCNMF: model based on collective matrix factorization.

Johnson, Alistair EW, et al. "MIMIC-III, a freely accessible critical care database." Scientific data 3 (2016): 160035. https://mimic.physionet.org/

Experiments and Results

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

Experimental Setup



Experiments and Results

DEPARTMENT OF COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

Diagnosis-Medication Correspondence

Table 1: Top Five Corresponding Medications for Three Diagnoses Inferred by HITF and Rubik

Cardiac dysrhythmias(39.0%)		Diabetes mellitus(25.3%)		Asthma(5.5%)	
HITF	**Rubik**	**HITF**	**Rubik**	**HITF**	**Rubik**
Furosemide(0.08)	Potassium Chloride(0.08)	Insulin(0.64)	Insulin(0.09)	Albuterol 0.083% Neb Soln(0.46)	Potassium Chloride(0.08)
Potassium Chloride(0.07)	Insulin(0.06)			Ipratropium Bromide Neb(0.39)	Insulin(0.06)
Metoprolol(0.06)	Furosemide(0.06)			Furosemide(0.08)	Furosemide(0.05)
Amiodarone HCl(0.05)	Magnesium Sulfate(0.04)	Furosemide(0.03)	Magnesium Sulfate(0.03)		
Heparin Sodium(0.04)	Acetaminophen(0.03)	Atorvastatin(0.03)	Acetaminophen(0.03)		

unrelated

Relevant drug identified by HITF gets much higher weight

Relevant drugs inferred only by HITF

Evaluated by a medical expert:

"There is qualitative superiority of HITF method over the Rubik method."
-- Medical expert

Experiments and Results

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

- Clinical relevance of the Phenotypes

Table 2: Three Examples of Phenotypes

	Medications	Diagnoses
Diabetes related disease	Diabetes mellitus Other diseases of lung Acute kidney failure Essential hypertension ...	Insulin Insulin Human Regular
Cardiac disease	Cardiac dysrhythmias Heart failure Other diseases of lung	Amiodarone HCl Metoprolol Furosemide ...
Respiratory disease	Other diseases of lung Cardiac dysrhythmias Heart failure Chronic airway obstruction, not elsewhere classified	Albuterol Diltiazem Ipratropium Bromide MDI Fluticasone Propionate ...

According to the medical expert, phenotypes inferred by HITF are clinically relevant.

Experiments and Results

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

Interpretability of the Phenotypes

Sparsity:

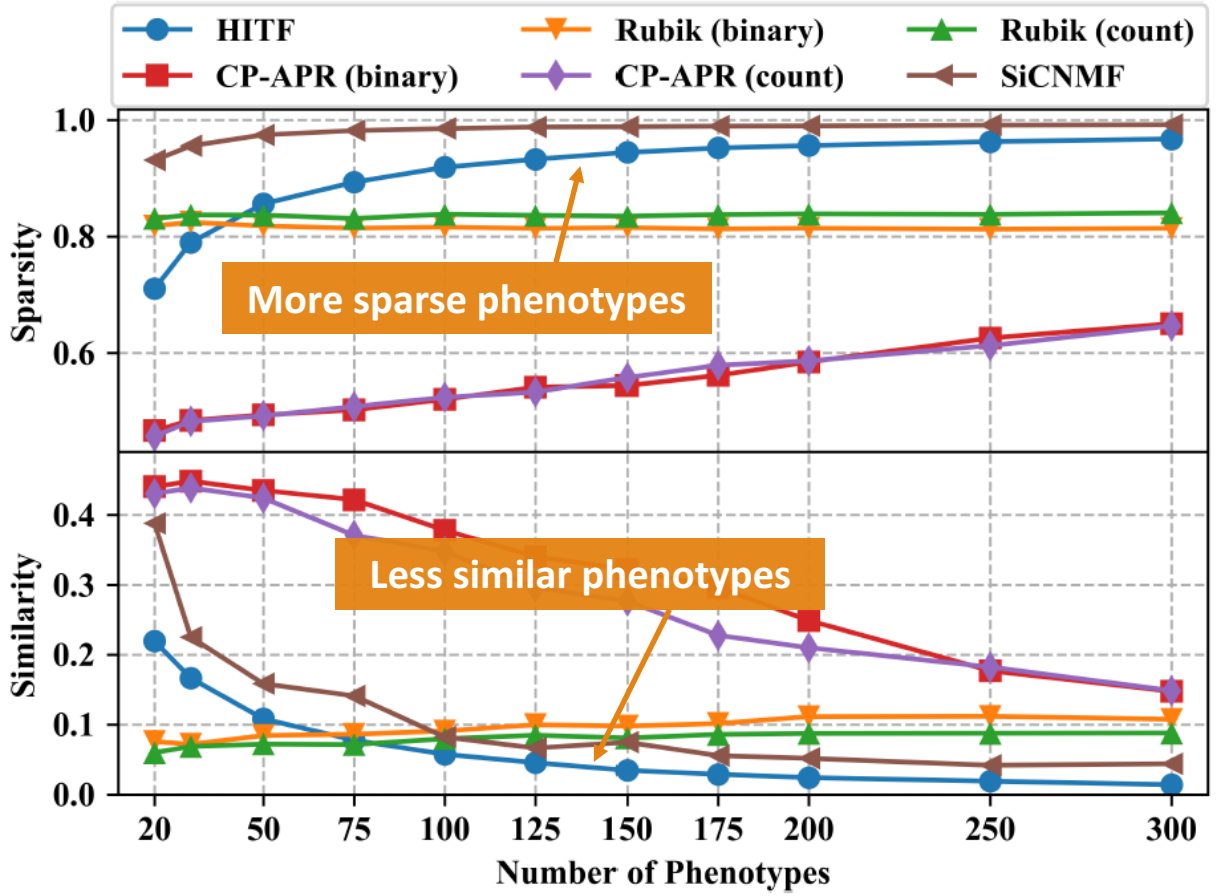
number of zero entries in one phenotype candidate.
The higher, the fewer elements in one phenotype.

Similarity:

average cosine similarity among all phenotypes.

$$\frac{\sum_{r_1}^{R}\sum_{r_2>r_1}^{R}\left\{\cos(\mathbf{U}^{(2)}_{:r_1},\mathbf{U}^{(2)}_{:r_2})+\cos(\mathbf{U}^{(3)}_{:r_1},\mathbf{U}^{(3)}_{:r_2})\right\}}{R(R-1)}$$

The smaller, the less similar the phenotypes are.

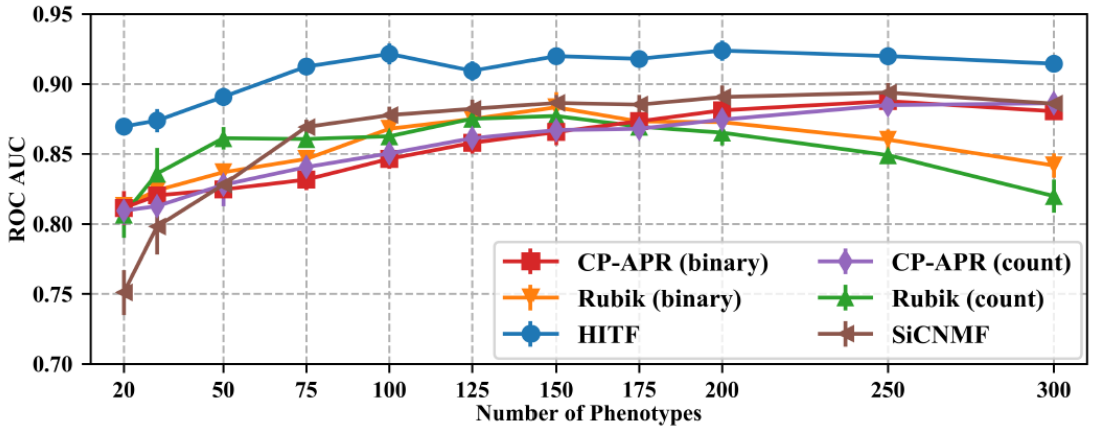


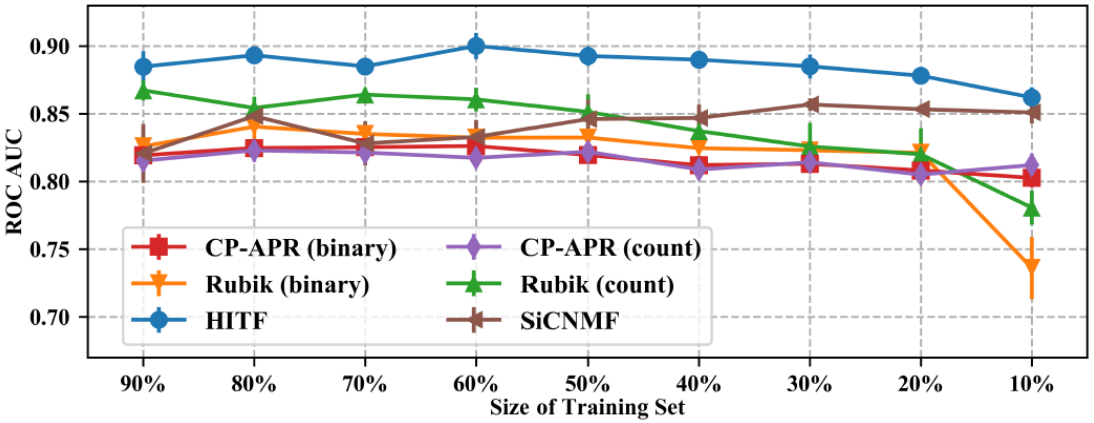
Phenotypes derived by HITF are highly interpretable.

Experiments and Results

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

Mortality prediction





- HITF outperforms all baselines consistently in terms of mortality prediction task.
- More robust against small size of training set.

Patients can be effectively represented by phenotypes derived using HITF.

Conclusion

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

- We proposed HITF to jointly learn the diagnosis-medication correspondence and phenotypes from EHR data simultaneously.
- Inferred diagnosis-medication correspondence is more reasonable and accurate than the “equal-correspondence” assumption.
- Phenotypes derived by HITF are clinical meaningful and interpretable.
- The predictive performance of HITF validates the effectiveness of representing patients using the derived phenotypes.

More information:
(codes will be released later)

香港浸會大學
HONG KONG BAPTIST UNIVERSITY

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

Thank you!

All questions and comments are greatly appreciated!

Yin Kejing
cskjyin@comp.hkbu.edu.hk