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

Kejing Yin¹, William K. Cheung¹, Yang Liu¹, Benjamin C. M. Fung², Jonathan Poon³

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

Background

- 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)
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.

---

Tensor Factorization as a Data-Driven Method

- Non-negative CP factorization for computational phenotyping:

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

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

Nature of tensor factorization: a low-rank model.

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

- Phenotype extraction from rank-one tensor:

<table>
<thead>
<tr>
<th>Diagnosis factor</th>
<th>Medication factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac dysrhythmias</td>
<td>Insulin 0</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>Metoprolol 0.55</td>
</tr>
<tr>
<td>Heart failure</td>
<td>Furosemide 0.45</td>
</tr>
<tr>
<td>Asthma</td>
<td>Albuterol 0</td>
</tr>
<tr>
<td>Pneumonitis</td>
<td>Diltiazem 0</td>
</tr>
</tbody>
</table>

Desired Properties of phenotypes for interpretability:
- More sparsity (less element in one phenotype)
- Less similarity (less similar between each other)


## Related Works

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Data distribution assumption and reconstruction error</th>
<th>Additional constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limestone</td>
<td>Poisson, KL divergence</td>
<td>No</td>
</tr>
<tr>
<td>Marble (KDD, 2014)</td>
<td>Poisson, KL divergence</td>
<td>Adding bias term (capturing global information)</td>
</tr>
<tr>
<td>Kim et al. (Scientific Report, 2017)</td>
<td>Gaussian, Frobenius norm</td>
<td>1. Prediction task (for improving discriminative power) 2. Clustering structure (making phenotypes distinct)</td>
</tr>
<tr>
<td>Kim et al. (KDD, 2017)</td>
<td>Gaussian, Frobenius norm</td>
<td>Distributed learning from multiple sites (without sharing patient level data)</td>
</tr>
</tbody>
</table>
Research Challenges

Interaction information are often missing in the records.

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

List of medications

<table>
<thead>
<tr>
<th>Medication</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vancomycin HCL</td>
<td>11</td>
</tr>
<tr>
<td>Metoprolol</td>
<td>14</td>
</tr>
<tr>
<td>Captopril</td>
<td>10</td>
</tr>
</tbody>
</table>

List of diagnoses

<table>
<thead>
<tr>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essential Hypertension</td>
</tr>
<tr>
<td>Pneumonitis</td>
</tr>
<tr>
<td>Type II Diabetes</td>
</tr>
</tbody>
</table>

Correspondence?
Unknown!

Construction of the tensor:

How to fill in the entries?

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

- Interaction information are often missing in the records.

Existing models adopt the “equal-correspondence” strategy:

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

Main Idea

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

Formulation

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

Sum of independent Poisson’s yields also a Poisson:

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

For diagnoses, we observe only a binarization:

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

Log-Likelihood:

$$\mathcal{L} = \mathcal{L}(M) + \mathcal{L}(D') = \sum_{i,k} \log \left( p(m_{ik} | U^{(n)}) \right) + \sum_{i,j} \log \left( p(d'_{ij} | U^{(n)}) \right)$$

$$= 1^T \left( -U^{(1)} \text{diag}(1^T U^{(2)}) U^{(3)^T} + M \ast \log(U^{(1)} \text{diag}(1^T U^{(2)}) U^{(3)^T}) \right) 1$$

$$+ 1^T \left( D' \ast \log(\exp(U^{(1)} \text{diag}(1^T U^{(3)}) U^{(2)^T}) - E) - U^{(1)} \text{diag}(1^T U^{(3)}) U^{(2)^T} \right) 1.$$

Infer the factor matrices by maximum likelihood estimation.
Learning Algorithms

Optimization Problem:

\[
\begin{align*}
\arg\min_{U^{(1)}, U^{(2)}, U^{(3)}} & \quad f(U^{(1)}, U^{(2)}, U^{(3)}) = -\mathcal{L}(M) - \mathcal{L}(D') \\
\text{subject to} & \quad U^{(n)} \geq 0, \quad n = 1, 2, 3.
\end{align*}
\]  

(13)

Apply block coordinate descent optimization:

Algorithm 1: Block Coordinate Descent Optimization Framework for HITF Model

Output: CP factor matrices: U^{(1)}, U^{(2)} and U^{(3)}

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

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

E.g., for the patient dimension: 

\[U^{(1)}_{k+1} = \arg\min_{X \geq 0} f(X, U^{(2)}_k, U^{(3)}_k),\]

Apply projected line search satisfying the Armijo condition:

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

Input: Variable X_k, search direction S_k, sufficient descent \(\sigma\) and descent step \(\rho\).
Output: Updated variable X_{k+1}

1. \(t \leftarrow 0;\)
2. while not \(f(P_+[X_k + \rho^i S_k]) - f(X_k) \leq \sigma ((P_+[X_k + \rho^i S_k] - X_k) \cdot \nabla f(X_k))\) do
   3. \(t \leftarrow t + 1;\)
3. end
4. update variable: \(X_{k+1} \leftarrow P_+[X_k + \rho^i S_k];\)

Take negative gradient as the search direction: 

\[S_k = -\nabla f(X_k)\]
Experiments and Results

- **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.

Experiments and Results

Experimental Setup

Training set

Testing set

Reconstructed interactions

Phenotypes

Patient representations (training set)

Logistic Regression Classifier

Mortality prediction

Experimental Setup

\[
\mathbf{X}' = \mathbf{M} \mathbf{D} \mathbf{D}'
\]

Training set

Testing set

Reconstructed interactions

Phenotypes

Patient representations (testing set)

Logistic Regression Classifier

Mortality prediction
Experiments and Results

Diagnosis-Medication Correspondence

<table>
<thead>
<tr>
<th>Cardiac dysrhythmias(39.0%)</th>
<th>Diabetes mellitus(25.3%)</th>
<th>Asthma(5.5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HITF</strong></td>
<td><strong>Rubik</strong></td>
<td><strong>HITF</strong></td>
</tr>
<tr>
<td>Furosemide(0.08)</td>
<td>Potassium Chloride(0.08)</td>
<td><strong>Insulin(0.64)</strong></td>
</tr>
<tr>
<td>Potassium Chloride(0.07)</td>
<td><strong>Insulin(0.06)</strong></td>
<td><strong>unrelated</strong></td>
</tr>
<tr>
<td>Metoprolol(0.06)</td>
<td>Furosemide(0.06)</td>
<td>Furosemide(0.06)</td>
</tr>
<tr>
<td>Amiodarone HCl(0.05)</td>
<td>Magnesium Sulfate(0.04)</td>
<td>Furosemide(0.03)</td>
</tr>
<tr>
<td>Heparin Sodium(0.04)</td>
<td>Acetaminophen(0.03)</td>
<td>Atorvastatin(0.03)</td>
</tr>
</tbody>
</table>

Evaluated by a medical expert:

“There is qualitative superiority of HITF method over the Rubik method.”

-- Medical expert
Experiments and Results

Clinical relevance of the Phenotypes

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

<table>
<thead>
<tr>
<th>Medications</th>
<th>Diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes mellitus</td>
<td>Insulin</td>
</tr>
<tr>
<td>Other diseases of lung</td>
<td>Insulin Human Regular</td>
</tr>
<tr>
<td>Acute kidney failure</td>
<td></td>
</tr>
<tr>
<td>Essential hypertension</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac dysrhythmias</td>
<td>Amiodarone HCl</td>
</tr>
<tr>
<td>Heart failure</td>
<td>Metoprolol</td>
</tr>
<tr>
<td>Other diseases of lung</td>
<td>Furosemide</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Other diseases of lung</td>
<td>Albuterol</td>
</tr>
<tr>
<td>Cardiac dysrhythmias</td>
<td>Diltiazem</td>
</tr>
<tr>
<td>Heart failure</td>
<td>Ipratropium Bromide MDI</td>
</tr>
<tr>
<td>Chronic airway obstruction, not elsewhere classified</td>
<td>Fluticasone Propionate</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Experiments and Results

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.

\[
\frac{\sum_{r_1}^{R} \sum_{r_2>r_1}^{R} \{\cos(U_{r_1}^{(2)}, U_{r_1}^{(2)}) + \cos(U_{r_1}^{(3)}, U_{r_2}^{(3)}) \}}{R(R-1)}
\]

The smaller, the less similar the phenotypes are.

Phenotypes derived by HITF are highly interpretable.
Experiments and Results

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

- 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)
Thank you!

All questions and comments are greatly appreciated!

Yin Kejing
cskjyin@comp.hkbu.edu.hk