SNeCT: Integrative cancer data analysis via large scale network constrained Tucker decomposition

Dongjin Choi and Lee Sael
**Motivation**

- **Q:** How can we characterize cancer patients?
  - **A:** The Cancer Genome Atlas (TCGA) Pan-Cancer data provide rich data across 12 tumor types.

John N. Weinstein *et al.* *Nat Genet* 45(10), 1113-1120 (2013) doi:10.1038/ng.2764

Motivation

- How can we provide integrated analysis for multi-dimensional data?
- Pan-Cancer12 data consist of multi-platform data

Motivation

- How can we build a combined model exploiting gene networks?
- Gene association networks provide gene similarity information

John N. Weinstein et al. *Nat Genet* 45(10), 1113-1120 (2013) doi:10.1038/ng.2764
Overview

- Introduction
- Problem definition
- Proposed method
- Experiments
- Conclusion
Tensor

- A tensor is a multi-dimensional array
- Pan-can12 data are represented as a 3-D tensor
Tensor Factorization

- Given a tensor, decompose the tensor into a core tensor and factor matrices whose product approximates the original tensor

**CP Decomposition**

\[ \mathbf{X} \approx \mathbf{A} \mathbf{G} \mathbf{B} \]

**Tucker Decomposition (HOSVD)**

\[ \mathbf{X} \approx \mathbf{A} \mathbf{G} \mathbf{B} \]
Overview

- Introduction
- Problem definition
- Proposed method
- Experiments
- Conclusion
Tucker Decomposition

- Tucker decomposition (Tucker, 1966)
  - Widely-used tensor factorization method
  - Given a tensor, Tucker decomposition factorizes the tensor into product of a core tensor and orthogonal factor matrices

\[ \mathbf{X} \approx \widehat{\mathbf{X}} = \mathbf{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \]

s.t. \( \mathbf{A}^T \mathbf{A} = \mathbf{B}^T \mathbf{B} = \mathbf{C}^T \mathbf{C} = \mathbf{I} \)

Elementwise,

\[ x_{ijk} \approx \mathbf{G} \times_1 a_i \times_2 b_j \times_3 c_k \]

- \( a_i \): \( i \)-th row of \( \mathbf{A} \)
- \( b_j \): \( j \)-th row of \( \mathbf{B} \)
- \( c_k \): \( k \)-th row of \( \mathbf{C} \)
Tucker Decomposition (cont.)

- Formal problem definition
  - Given a 3-D tensor \( \mathcal{X} \in \mathbb{R}^{I \times J \times K} \) with observable entries \( \{x_{ijk} | (i, j, k) \in \Omega_{\mathcal{X}} \} \), the rank-\([P, Q, R]\) factorization of \( \mathcal{X} \) is to find the core tensor \( \mathcal{G} \) and factor matrices \( \{A, B, C\} \) which minimizes the following loss function:

\[
\begin{align*}
  f(\mathcal{G}, A, B, C) &= \frac{1}{2} \left\| \mathcal{X} - \hat{\mathcal{X}} \right\|_F^2 + \frac{\lambda}{2} R(\mathcal{G}, A, B, C) \\
  &= \frac{1}{2} \sum_{(i,j,k) \in \Omega_{\mathcal{X}}} \left(x_{ijk} - \mathcal{G} \times_1 a_i \times_2 b_j \times_3 c_k\right)^2 + \frac{\lambda}{2} R(\mathcal{G}, A, B, C)
\end{align*}
\]
Overview

- Introduction
- Problem definition
- Proposed method
- Experiments
- Conclusion
Scheme of SNeCT

**Input**

- Patients profile
- Gene
- Platform

**Lock-Free Parallel SGD**

- A
- B
- C
- g

**Extract patients profile**

- A
- B
- C

**Stratification**

- A
- C₁
- C₂
- Patients clustering

**Prediction**

- Query patient data
- \( a_q \)
- Top-k search

**Personalized Subtype Analysis**

- \( g \) x \( a_i \) = S
Proposed methods

- **SNeCT** enables integrative tensor factorization and analysis for tensor data with network constraint.
  
  \[ \text{SNeCT} = \text{Scalable Network Constrained Tucker decomposition} \]

- **Method 1**
  - Formulate SGD-amenable objective function
  - Iterative SGD update with lock-free parallel scheme

- **Method 2**
  - Personalized subtype analysis
Proposed methods

- Formulate SGD-amenable objective function
  - Given the gene similarity matrix $Y (\in \mathbb{R}^{J \times J})$ with observable entries $\{y_{mn}|(m, n) \in \Omega_Y\}$, network constraint is formulated to make similar genes have similar factors:

$$
\begin{align*}
    f_g (B, Y) &= \frac{1}{2} \sum_{l=1}^{Q} \left[ \sum_{(m,n) \in \Omega_Y} y_{mn} (b_{ml} - b_{nl})^2 \right] \\
    &= \frac{1}{2} \sum_{(m,n) \in \Omega_Y} y_{mn} \| b_m - b_n \|_F^2
\end{align*}
$$
Proposed methods

- Formulate SGD-amenable objective function

\[
f(G,A,B,C) = \frac{1}{2} \sum_{(i,j,k) \in \Omega_X} (x_{ijk} - \tilde{x}_{ijk})^2 + \frac{\lambda}{2} R(G,A,B,C)
\]

\[
= \frac{1}{2} \sum_{(i,j,k) \in \Omega_X} \left[ (x_{ijk} - \tilde{x}_{ijk})^2 + \frac{\lambda}{|\Omega_X|} \|G\|_F^2 + \lambda \left( \frac{\|a_i\|_F^2}{|\Omega_X^i|} + \frac{\|b_j\|_F^2}{|\Omega_X^j|} + \frac{\|c_k\|_F^2}{|\Omega_X^k|} \right) \right]
\]

\[
f_g (B,Y) = \frac{1}{2} \sum_{(m,n) \in \Omega_Y} y_{mn} \|b_m - b_n\|_F^2
\]

- Integrate into single objective function

\[
f_{opt} = f + \lambda_g f_g
\]
Proposed methods

- Calculate gradients of $f_{opt}$ with respect to the core tensor and factor matrices for a given data point $x_\alpha= (ijk)$ or $y_\beta= (mn)$

\[
\frac{\partial f_{opt}}{\partial a_i} \bigg|_\alpha = - (x_\alpha - \bar{x}_\alpha)[G \times_2 b_j \times_3 c_k] + \frac{\lambda}{|\Omega_x|} a_i
\]

\[
\frac{\partial f_{opt}}{\partial G} \bigg|_\alpha = - (x_\alpha - \bar{x}_\alpha) \times_1 a_i^T \times_2 b_j^T \times_3 c_k^T + \frac{\lambda}{|\Omega_x|} G
\]

\[
\frac{\partial f_{opt}}{\partial b_m} \bigg|_\beta = \lambda_g y_\beta (b_m - b_n)
\]

- $\frac{\partial f_{opt}}{\partial b_j} \bigg|_\alpha$, $\frac{\partial f_{opt}}{\partial c_k} \bigg|_\alpha$, and $\frac{\partial f_{opt}}{\partial b_n} \bigg|_\beta$ are calculated symmetrically.
Proposed methods

- Parallel update with calculated gradient

- SNeCT(\(\mathcal{X}, \mathcal{Y}, \lambda, \lambda_g, \eta\)) (\(\eta\): learning rate)

1. Initialize \(G, A, B, C\) randomly
2. repeat
3. for \(\forall x_{ijk} = \alpha \in \mathcal{X}, \forall y_{mn} = \beta \in \mathcal{Y}\) in random order in parallel
4. if \(x_{ijk} \in \mathcal{X}\) is picked then
5. \(a_i \leftarrow a_i - \eta \frac{\partial f_{opt}}{\partial a_i} |_{\alpha}\), \(b_j \leftarrow b_j - \eta \frac{\partial f_{opt}}{\partial b_j} |_{\alpha}\), \(c_k \leftarrow c_k - \eta \frac{\partial f_{opt}}{\partial c_k} |_{\alpha}\)
6. \(G \leftarrow G - \eta \frac{\partial f_{opt}}{\partial G} |_{\alpha}\)
7. else if \(\forall y_{mn} \in \mathcal{Y}\) is picked then
8. \(b_m \leftarrow b_m - \eta \frac{\partial f_{opt}}{\partial b_m} |_{\beta}\), \(b_n \leftarrow b_n - \eta \frac{\partial f_{opt}}{\partial b_n} |_{\beta}\)
9. end if
10. end for
11. until convergence condition satisfied
12. Orthogonalize \(A, B, C\) by QR decomposition
13. return \(G, A, B, C\)
Overview

- Introduction
- Problem definition
- Proposed method
- Experiments
- Conclusion
Experimental Settings

- Factorize data tensor with rank-\([78,48,5]\)
- **Stratification**
  - Cluster analysis
  - Survival analysis
- **Prediction**
  - Top-k similarity search on clinical features
- **Personalized subtype analysis**
- **Performance**
  - Compare speed and convergence rate with competitor
  - Competitor: Narita et al. 2012
## Stratification – Cluster Analysis

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>C9</th>
<th>C10</th>
<th>C11</th>
<th>C12</th>
<th>C13</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLCA</td>
<td>16</td>
<td>32</td>
<td>2</td>
<td>19</td>
<td>0</td>
<td>22</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>126</td>
</tr>
<tr>
<td>BRCA</td>
<td>17</td>
<td>3</td>
<td>600</td>
<td>172</td>
<td>1</td>
<td>70</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>889</td>
</tr>
<tr>
<td>COAD</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>91</td>
<td>317</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>419</td>
</tr>
<tr>
<td>GBM</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>248</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>267</td>
</tr>
<tr>
<td>HNSC</td>
<td>0</td>
<td>242</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>310</td>
</tr>
<tr>
<td>KIRC</td>
<td>14</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>471</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>498</td>
</tr>
<tr>
<td>LAML</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>188</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>197</td>
</tr>
<tr>
<td>LUAD</td>
<td>302</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>457</td>
</tr>
<tr>
<td>LUSC</td>
<td>26</td>
<td>32</td>
<td>0</td>
<td>29</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>246</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>340</td>
</tr>
<tr>
<td>OV</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>348</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>131</td>
</tr>
<tr>
<td>READ</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>9</td>
<td>145</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>163</td>
</tr>
<tr>
<td>UCEC</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>117</td>
<td>1</td>
<td>348</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>13</td>
<td>2</td>
<td>0</td>
<td>499</td>
</tr>
<tr>
<td>Total</td>
<td>387</td>
<td>315</td>
<td>613</td>
<td>362</td>
<td>477</td>
<td>581</td>
<td>467</td>
<td>348</td>
<td>249</td>
<td>188</td>
<td>412</td>
<td>17</td>
<td>134</td>
<td>4550</td>
</tr>
</tbody>
</table>
Stratification – Survival Analysis

- Survival curves for clustered patients

---

Introduction
Problem definition
Proposed method
Experiments
Conclusion

log-rank statistics:

- A: \( \lambda = 0 \) with 409 patients
- B: \( \lambda = 0.1 \) with 1151 patients
- C: \( \lambda = 1 \) with 1185 patients

Days after diagnosis

Survival ratio

C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13
Prediction – Top-k similarity search

- When a new query patient \( q \) arrives with data \( X_q \), calculate factor \( a_q \) satisfying following equation: \( a_q = \arg \min_a \| X_q - G \times_a 1 \times B \times C \| \)

- Find top-k similar patients to \( q \) and compare

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Clinical Features</th>
<th>Top 1</th>
<th>Top 5</th>
<th>Top 10</th>
<th>Top R</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCA</td>
<td>Estrogen receptor status</td>
<td>0.72</td>
<td>0.85</td>
<td>0.86</td>
<td>0.81</td>
</tr>
<tr>
<td>COAD</td>
<td>Braf gene analysis result</td>
<td>1.00</td>
<td>0.80</td>
<td>0.70</td>
<td>0.92</td>
</tr>
<tr>
<td>GBM</td>
<td>Histological type</td>
<td>0.96</td>
<td>0.94</td>
<td>0.94</td>
<td>0.78</td>
</tr>
<tr>
<td>HNSC</td>
<td>Hpv status by p16 testing</td>
<td>0.78</td>
<td>0.78</td>
<td>0.77</td>
<td>0.73</td>
</tr>
<tr>
<td>KIRC</td>
<td>Histological type</td>
<td>1.00</td>
<td>0.99</td>
<td>0.99</td>
<td>0.73</td>
</tr>
<tr>
<td>LAML</td>
<td>Calgb cytogenetics risk cat.</td>
<td>0.85</td>
<td>0.84</td>
<td>0.81</td>
<td>0.65</td>
</tr>
<tr>
<td>OV</td>
<td>Neoplasm histologic grade</td>
<td>0.79</td>
<td>0.75</td>
<td>0.76</td>
<td>0.77</td>
</tr>
<tr>
<td>READ</td>
<td>Braf gene analysis result</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>UCEC</td>
<td>Menopause status</td>
<td>0.71</td>
<td>0.76</td>
<td>0.76</td>
<td>0.77</td>
</tr>
</tbody>
</table>
Personalized subtype analysis

- To provide personalized interpretation for patient \( i \), calculate \( G \times_1 a_i = S(\in R^{Q \times R}) \)
- Norms of rows represent gene subtype influence
- Norms of columns represent platform subtype influence
Performance

- Comparison with another network-constrained tensor factorization method: Narita et al. 2012
  - **A. Speed**: Iteration time – measured on sampled data
  - **B. Accuracy**: Test RMSE
Overview

- Introduction
- Problem definition
- Proposed method
- Experiments
- Conclusion
Conclusion

- SNeCT
  - Parallel algorithms for network constrained tensor factorization
  - Solve tucker decomposition through parallel SGD update scheme
  - Engage common pathway gene network into Pan-Caner12 tensor
  - Utilize patient factor matrix on cluster analysis and survival analysis
  - Propose a personalized subtype analysis scenario
Thank you!

Questions?