Cox Regression with Correlation based Regularization for Electronic Health Records

Bhanukiran Vinzamuri, Chandan K. Reddy

bhanukiranj@wayne.edu, reddy@cs.wayne.edu
Department of Computer Science
Wayne State University, Detroit Michigan USA

10 December 2013
Outline of the presentation

Mining EHR data

Survival Regression

Cox Regression with Correlation Regularization

Experimental Results

Conclusions and Future Work
Electronic Health Records (EHR) data consist of diverse patient measurements concatenated from different sources. Diverse sources such as Labs, Claims and Demographics make mining EHR data challenging! Problems of societal impact such as Heart Failure Readmission Risk prediction are determined from EHR.
Challenges

- Feature construction: Pruning redundant features and creating new features of importance.
- Correlated features: Handling correlated and non linear features in EHR datasets is important when building regression models.
- Survival probability: Estimating the survival probability for a patient at a given time after discharge.
Survival Regression

- Survival Regression deals with building predictive models for time to event data. Time to event data is defined based on an event such as heart failure readmission or death of a patient.
- These methods can predict the survival probability for a patient using the notion of a hazard function and a linear regression formulation.
- The hazard function models the probability of occurrence of the event (readmission/death) for any given patient at a time point.
Censoring in survival data

Figure: Censoring in survival data
Cox Regression

- A method for predicting the survival probability in censored data.
- Semi parametric method of regression coefficient estimation.

\[
h(t) = h_0(t) \exp(\sum_{j=1}^{n} x_j^T \beta) \quad (1)
\]

- Simple formulation has lead to widespread use among biostatistics researchers and clinical community.
Re-formulation of Cox Regression as WLSQ problem

\[ \hat{\beta} = \min_{\beta} \frac{1}{n} \sum_{k=1}^{n} (W(k, k)((Z(k, :) - X(k, :)\beta)^2) \quad (2) \]

- \( \beta \) is the regression coefficient vector.
- \( W \) is the symmetric weight matrix.
- \( Z \) is the pseudo response vector.
Cox Regression as WLSQ problem

\[ W(k, k) = \sum_{i \in C_k} \left[ \frac{e^{\tilde{\eta}(k,:)} \sum_{j \in R_i} e^{\tilde{\eta}(j,:)} - (e^{\tilde{\eta}(k,:)} )^2}{(\sum_{j \in R_i} e^{\tilde{\eta}(j,:)} )^2} \right] \]  \hspace{1cm} (3)

\[ Z(k,:) = \tilde{\eta}(k,:) + \frac{1}{W(k,k)} \left[ \delta_k - \sum_{i \in C_k} \left( \frac{e^{\tilde{\eta}(k,:)} }{ \sum_{j \in R_i} e^{\tilde{\eta}(j,:)}} \right) \right] \]  \hspace{1cm} (4)

- \( R_i \) is the set of all patients \( j \) at risk at time \( t_i \) (\( t_j > t_i \)).
- \( C_k \) is the set of all times for which patient \( k \) is still at risk. (\( t_i < t_k \))
Survival Function estimation in Cox Regression

\[ h_0(t) = \sum_{t_i \leq t} \frac{\delta_i}{\sum_{j \in R_i} \exp(\hat{\beta}^T x_j)} \]  \hspace{1cm} (5)

\[ s_0(t) = \exp(-h_0(t)) \]

\[ s(t|x_i) = s_0(t)\exp(\hat{\beta}^T x_i) \]

- \( h_0(t) \) is the baseline hazard function.
- \( s_0(t) \) is the baseline survival function.
- \( s(t|x_i) \) is the conditional survival probability for patient \( i \).
Necessity of Correlation based Regularization in Cox Regression

- Regularization to handle \textit{correlated} attributes in EHR data. (KEN-COX)
- Regularization to handle grouped \textit{correlated} features and structured sparsity in EHR data. (OSCAR-COX)
Kernel Elastic Net (KEN) Formulation

\[
y = X\beta + \epsilon \quad (6)
\]

\[
1^T y = 0 \quad 1^T X_i = 0 \quad X_i^T X_i = 1; \quad (7)
\]

\[
\hat{\beta} = \arg \min_{\beta} ||y - X\beta||^2 + \eta J(\beta) \quad (8)
\]

\[
J(\beta) = |\beta|^T K |\beta| \quad (9)
\]

\[
\hat{\beta} = \arg \min_{\beta} ||y - X\beta||^2 + \eta |\beta|^T K |\beta|
\]
Kernel Elastic Net Cox Regression (KEN-COX)

- Use kernel elastic net penalty in the re-formulated weighted linear regression framework of Cox Regression.
- Cyclic coordinate descent based optimization algorithm.
- Estimate the $p^{th}$ coordinate keeping all the remaining ones constant. This is done iteratively until convergence!
- Update step is as follows

$$\hat{\beta}(p, :) = \frac{S(\frac{1}{n} \sum_{k=1}^{n} W(k,k)X(k,p)[Z(k,:) - \sum_{j \neq p} X(k,j)\tilde{\beta}(j,:)],\lambda\alpha)}{\frac{1}{n} \sum_{k=1}^{n} W(k,k)X(k,p)^2 + \lambda\alpha K(p,p)}$$
Grouping of features in EHR data

- Feature grouping improves the stability of feature selection, and it can help in gaining additional insight.
- KEN-COX can handle correlation, but it cannot group multiple correlated features in high dimensional EHR data!
- Need to discover groups automatically from the data without the need to pre specify grouping structure.
The first part of the OSCAR regularizer encourages sparsity as in lasso.

The second part is a pairwise $L_\infty$ regularizer which encourages every coefficient pairs $|\beta_i|, |\beta_j|$ to be equal.

$$\hat{\beta} = \arg \min_{\beta} \| y - X\beta \|^2 + \lambda_1 \| \beta \|_1 + \lambda_2 \sum_{i<j} \max\{|\beta_i|, |\beta_j|\} \quad (10)$$
ADMM optimization of OSCAR-COX

- Alternate Direction Method of Multipliers is fast at converging to optimal solution!
- $L$ is the cox partial log likelihood loss function. $T$ is obtained from the attribute based graph representation of the dataset.

\[
\hat{\beta} = \min_{\beta} L(\beta) + \lambda_1(\| \beta \|_1) + \lambda_2(\| T\beta \|_1) \tag{11}
\]

\[
\hat{\beta} = \min_{\beta, q, p} L(\beta) + \lambda_1 \| q \|_1 + \lambda_2 \| p \|_1 \\
\text{s. t. } \beta - q = 0, \quad T\beta - p = 0 \tag{12}
\]
OSCAR based Cox Regression (OSCAR-COX)

\[ L_\rho(\beta, q, p) = L(\beta) + \lambda_1 \| q \|_1 + \lambda_2 \| p \|_1 \]
\[ + \mu^T (\beta - q) + \nu^T (T\beta - p) \]
\[ + \frac{\rho}{2} \| \beta - q \|^2 + \frac{\rho}{2} \| T\beta - p \|^2 \tag{13} \]

\[ \hat{\beta} = \min_\beta L_\rho(\beta, \hat{q}, \hat{p}) \tag{14} \]

\[ \hat{q} = \min_q L_\rho(\hat{\beta}, q, \hat{p}) \tag{15} \]

\[ \hat{p} = \min_p L_\rho(\hat{\beta}, \hat{q}, p) \tag{16} \]
Experimental Setup

- 18,701 EHR instances for 8,692 patients admitted for heart failure over a duration of 10 years were considered.
- Individual longitudinal EHR datasets (DS1-DS5) were generated from these instances.
- Synthetic datasets were generated from a $N(0, 1)$ and survival times were generated using a Weibull distribution. Parameters were modified to generate 5 synthetic datasets Syn1-Syn5.
Datasets description

<table>
<thead>
<tr>
<th>Dataset</th>
<th># Features</th>
<th># Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syn1</td>
<td>15</td>
<td>5000</td>
</tr>
<tr>
<td>Syn2</td>
<td>50</td>
<td>5000</td>
</tr>
<tr>
<td>Syn3</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Syn4</td>
<td>1000</td>
<td>100</td>
</tr>
<tr>
<td>Syn5</td>
<td>500</td>
<td>50</td>
</tr>
<tr>
<td>DS1</td>
<td>732</td>
<td>5675</td>
</tr>
<tr>
<td>DS2</td>
<td>709</td>
<td>4379</td>
</tr>
<tr>
<td>DS3</td>
<td>668</td>
<td>3543</td>
</tr>
<tr>
<td>DS4</td>
<td>658</td>
<td>2826</td>
</tr>
<tr>
<td>DS5</td>
<td>609</td>
<td>2278</td>
</tr>
</tbody>
</table>
**Goodness of fit**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>OSCAR-COX</th>
<th>KEN-COX</th>
<th>EN-COX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>$R^2$</td>
<td>MSE</td>
</tr>
<tr>
<td>DS1</td>
<td>2.85</td>
<td>0.41</td>
<td>2.96</td>
</tr>
<tr>
<td>DS2</td>
<td>2.78</td>
<td>0.40</td>
<td>2.97</td>
</tr>
<tr>
<td>DS3</td>
<td>2.27</td>
<td>0.35</td>
<td>2.95</td>
</tr>
<tr>
<td>DS4</td>
<td>2.03</td>
<td>0.48</td>
<td>2.79</td>
</tr>
<tr>
<td>DS5</td>
<td>2.9</td>
<td>0.29</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Table: Comparison of MSE and $R^2$ values of KEN-COX and OSCAR-COX with EN-COX and COX
Non Redundant Feature Selection

Table: Comparison in the redundancy of features using Redundancy metric

<table>
<thead>
<tr>
<th>Method</th>
<th>DS1</th>
<th>DS2</th>
<th>DS3</th>
<th>DS4</th>
<th>DS5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relief-F</td>
<td>0.05</td>
<td><strong>0.039</strong></td>
<td>0.07</td>
<td>0.055</td>
<td>0.06</td>
</tr>
<tr>
<td>Fisher Score</td>
<td>0.042</td>
<td>0.043</td>
<td><strong>0.05</strong></td>
<td>0.067</td>
<td>0.07</td>
</tr>
<tr>
<td>SPEC</td>
<td>0.042</td>
<td>0.045</td>
<td><strong>0.05</strong></td>
<td>0.054</td>
<td>0.06</td>
</tr>
<tr>
<td>mRmR</td>
<td>0.04</td>
<td>0.046</td>
<td>0.054</td>
<td>0.051</td>
<td>0.058</td>
</tr>
<tr>
<td>KEN-COX</td>
<td><strong>0.039</strong></td>
<td>0.047</td>
<td>0.051</td>
<td>0.048</td>
<td>0.052</td>
</tr>
<tr>
<td>OSCAR-COX</td>
<td>0.042</td>
<td>0.045</td>
<td>0.052</td>
<td><strong>0.043</strong></td>
<td><strong>0.045</strong></td>
</tr>
</tbody>
</table>

Redundancy = \( \frac{1}{m(m-1)} \sum_{f_i,f_j \in F_i > j} \rho_{ij} \) \hspace{1cm} (17)
**Comparison with Linear and Logistic Learners**

Table: Comparison of AUC values for KEN-COX and OSCAR-COX against regularized linear and logistic learners

<table>
<thead>
<tr>
<th>Dataset</th>
<th>OSCAR-COX</th>
<th>KEN-COX</th>
<th>LASSO</th>
<th>EN</th>
<th>FLA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syn1</td>
<td>0.81</td>
<td><strong>0.8472</strong></td>
<td>0.64</td>
<td>0.68</td>
<td>0.80</td>
</tr>
<tr>
<td>Syn2</td>
<td>0.8814</td>
<td>0.8605</td>
<td>0.76</td>
<td>0.784</td>
<td><strong>0.8918</strong></td>
</tr>
<tr>
<td>Syn3</td>
<td><strong>0.8909</strong></td>
<td>0.8412</td>
<td>0.76</td>
<td>0.762</td>
<td>0.874</td>
</tr>
<tr>
<td>Syn4</td>
<td><strong>0.888</strong></td>
<td>0.8605</td>
<td>0.72</td>
<td>0.69</td>
<td>0.8533</td>
</tr>
<tr>
<td>Syn5</td>
<td><strong>0.8875</strong></td>
<td>0.859</td>
<td>0.71</td>
<td>0.75</td>
<td>0.8575</td>
</tr>
</tbody>
</table>
Biomarker identification from Heart failure EHR data

**Table:** Association between biomarkers and heart failure readmission

<table>
<thead>
<tr>
<th>Variable</th>
<th>Assoc</th>
<th>LASSO</th>
<th>COX</th>
<th>OSCAR-COX</th>
<th>KEN-COX</th>
</tr>
</thead>
<tbody>
<tr>
<td>HGB</td>
<td>0.81</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ckd</td>
<td>0.75</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>diabetes</td>
<td>0.71</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>hypertension</td>
<td>0.70</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BUN/CREAT</td>
<td>0.66</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>age</td>
<td>0.66</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>cad</td>
<td>0.61</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>heart_failure</td>
<td>0.60</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>afib</td>
<td>0.60</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Conclusions and Future Work

- We developed a novel (KEN-COX) regression algorithm which is more effective than elastic net regularized Cox (EN-COX) at handling correlated features.
- We developed a novel OSCAR-COX regression algorithm which captures group correlation and structured sparsity in the EHR data effectively.
- We proposed scalable and efficient optimization routines for both these algorithms.
- Future work includes integrating multi task regularized learning algorithms into cox regression.
Questions

Thank you!