Medical Coding Classification by Leveraging Inter-Code Relationships

Y. Yan+, G. Fung*, J. Dy+, R. Rosales*

+Northeastern University, *Siemens Healthcare

July 18th 2010
Medical Coding

-is the process of transforming information contained in patient medical records into standard pre-defined medical codes.

Sample Medical Transcript

Description: A 23-year-old white female presents with complaint of allergies.

SUBJECTIVE: This 23-year-old white female presents with complaint of allergies. She used to have allergies when she lived in Seattle but she thinks they are worse here. In the past, she has tried Claritin and Zyrtec. Both worked for short time but then seemed to lose effectiveness. She has used Allegra also. She used that last summer and she began using it again two weeks ago. It does not appear to be working very well. She has used over-the-counter sprays but no prescription nasal sprays. She does have asthma but does not require daily medication for this and does not think it is flaring up.

MEDICATIONS: Her only medication currently is Ortho Tri-Cycle and the Allegra.

ALLERGIES: She has no known medicine allergies.

OBJECTIVE:
Vitals: Weight was 130 pounds and blood pressure 128/78.
HEENT: Her throat was mildly erythematous without exudate. Nasal mucosa was erythematous and swollen. Only clear drainage was seen. TMS were clear.
Neck: Supple without adenopathy.
Lungs: Clear.

ASSESSMENT: Allergic rhinitis.

PLAN:
1. She will try Zyrtec instead of Allegra again. Another option will be to use loratadine. She does not think she has prescription coverage so that might be cheaper.
2. Samples of Nasonex two sprays in each nostril given for three weeks. A prescription was written as well.

ICD-9, the Ninth Revision of ICD (International Statistical Classification of Diseases and Related Health Problems), provides the standard for coding clinical records in the US.

- The codes include diagnosis information, classifications for signs, symptoms, abnormal findings, complaints, social circumstances, and causes of injury or disease.
- Useful for clinical care, research, education and billing purposes.

Example Codes:

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>493.1</td>
<td>Intrinsic asthma</td>
</tr>
<tr>
<td>786.2</td>
<td>Cough</td>
</tr>
<tr>
<td>98.83</td>
<td>Domestic tasks therapy</td>
</tr>
<tr>
<td>970.1</td>
<td>Nalorphine</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Characteristics for Problem

- Sparsity of Classes
- High Dimensionality
- Multi-Label Classification Problem
Typical Approaches to Multi-Label Classification

- 1-vs-Rest
  - 1-vs-Rest SVM
  - K-Nearest Neighbors

- Voting
  - \textit{rank}-SVM

- Probabilistic Modeling
  - BPMLL
Formulation

**Objective Function**

Proximal SVM with $l1$-norm

$$\min_{W,M} \mu \sum_{i=1}^{D} \|e - \text{diag}(y_i)\hat{y}_i\|^2 + \|W\|_1$$

where

$\hat{y}_i \triangleq (x_i W I + \gamma)$

**Notations**

- Inputs: $x$, $X$
- Labels: $y$, $Y$
- Classifiers: $W$
- Relation Matrix: $M$
- Prior: $\tilde{M}_{\text{sim}}$
- Shifts: $\gamma$
- One Vector: $e$
- Trade Off: $\mu$, $\nu$
Formulation

Objective Function

**Proximal SVM with l1-norm**

\[
\min_{W, M} \mu \sum_{i=1}^{D} \| e - \text{diag}(y_i) \hat{y}_i' \|^2 + \| W \|_1 + \nu \| M - \tilde{M}_{\text{sim}} \|_{\text{frob}}^2
\]

\[\text{s.t.} \quad -1 \leq M(i, j) \leq 1 \quad i, j = \{1, 2, \ldots, D\}\]

where

\[\hat{y}_i \triangleq (x_i WM + \gamma)\]

**Regularization on Class Relationship Matrix**

Notations

**Inputs:** \( x, X \)  
**Labels:** \( y, Y \)  
**Classifiers:** \( W \)  
**Relation Matrix:** \( M \)  
**Prior:** \( \tilde{M}_{\text{sim}} \)  
**Shifts:** \( \gamma \)  
**One Vector:** \( e \)  
**Trade Off:** \( \mu, \nu \)
Class Relationship Matrix Prior, $\tilde{M}_{\text{sim}}$

$$M_{\text{sim}} = [\omega_{ij}]_{L \times L}$$

$$\omega_{ij} = \sum_{q=1}^{D} y_q(i)y_q(j) \leftarrow \text{Similarity between Classes (Codes)}$$

$$D_M = [d_{ij}]_{L \times L} \leftarrow \text{Degree Matrix}$$

$$d_{ii} = \sum_{j} |\omega_{ij}| \quad d_{ij} = 0 \ (i \neq j)$$

$$\tilde{M}_{\text{sim}} = M_{\text{sim}}D_{M}^{-1}$$

<table>
<thead>
<tr>
<th>samples</th>
<th>word-1</th>
<th>word-2</th>
<th>...</th>
<th>class</th>
<th>code-1</th>
<th>code-2</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>-1.2</td>
<td>231</td>
<td>...</td>
<td>$y_1$</td>
<td>-1</td>
<td>-1</td>
<td>...</td>
</tr>
<tr>
<td>$x_2$</td>
<td>2.4</td>
<td>50</td>
<td>...</td>
<td>$y_2$</td>
<td>-1</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>$x_3$</td>
<td>0.5</td>
<td>873</td>
<td>...</td>
<td>$y_3$</td>
<td>-1</td>
<td>-1</td>
<td>...</td>
</tr>
</tbody>
</table>
Large Scale Formulation

- For large scale problems, it is not computationally efficient to solve the original form of the objective function.
- We re-write the objective function in vector form (factorize both $W$ and $M$), and update them one vector at a time.

$$
\min_{W,M} \mu \sum_{i=1}^{D} \|y_i - x_i \sum_{j=1}^{L} w_j m_{j}^{r'}\|^2 + \sum_{j=1}^{L} \|w_j\|_1 + \nu \|M - \tilde{M}_{sim}\|_{frob}^2
$$

s.t. $-1 \leq M(i,j) \leq 1 \quad i,j = \{1, 2, \ldots, D\}$

where:

$$
W = [w_1, w_2, \cdots, w_L]; \quad M = [m_1^{r'}, m_2^{r'}, \cdots, m_L^{r'}]'
$$
Experiment Results

ACC. of ICD-9

Average ROC of ICD-9
R.O.C for Different Msim

ROC for Different $M_{sim}$

ROC: effect of different initial Msim (ICD9)

- **Similarity**: AUC: 0.96
- **Identity**: AUC: 0.92
- **Random**: AUC: 0.89

FPR or (1−specificity) vs TPR or sensitivity
Learned Relations between Classes (Codes)
## Feature Selected

**Feature with the Largest ABS. Value (Up to 5) per Class**

<table>
<thead>
<tr>
<th>class 1</th>
<th>class 2</th>
<th>class 3</th>
<th>class 4</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pneu-</td>
<td>Asthma</td>
<td>Pulm.</td>
<td>Rheu.</td>
<td></td>
</tr>
<tr>
<td>postvoid</td>
<td>lobe</td>
<td>atelectasis</td>
<td>hydron-</td>
<td></td>
</tr>
<tr>
<td>brochial</td>
<td>pneumonia</td>
<td>collapse</td>
<td>ephrosis</td>
<td></td>
</tr>
<tr>
<td>ultrasound</td>
<td>wheezing</td>
<td></td>
<td>cough</td>
<td></td>
</tr>
<tr>
<td>pneumonia</td>
<td>asthma</td>
<td></td>
<td>atelectasis</td>
<td></td>
</tr>
<tr>
<td>cystitis</td>
<td>opacity</td>
<td></td>
<td>pyejectasis</td>
<td></td>
</tr>
<tr>
<td>pneumonia</td>
<td>asthma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cystitis</td>
<td>opacity</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary

Introduced a novel Multi-Label approach that...

1. Makes use of Inter-Class relations and learns the relations
2. Can work even when the samples in each class are sparse
3. Can perform feature selection