**Forecasting Potential Diabetes Complications**

*Yang Yang, Walter Luyten, Lu Liu, Marie-Francine Moens, Jie Tang, Juanzi Li*
Tsinghua University, Katholieke Universiteit Leuven, Northwestern University

- Diabetes are major causes of early death in most countries and over 68% of diabetes-related mortality is caused by diabetes complications.
- 471 billion USD were spent on healthcare for 371 million diabetes patients world-wide in 2012, still 4.8 million people died in 2012 due to diabetes.

**Proposed Model**

**Key challenge: feature sparseness**
- Averagely each clinical record only contains 1.26% of lab tests on average.
- 65.5% types of lab tests are recorded in less than 0.0054% of clinical records.

The model alleviates the sparseness issue by projecting feature space into a low-dimensional latent space.

**Forecasting Performance**
- Dataset: a collection of real medical records from a famous geriatric hospital:
  - 181,933 medical records, 35,525 patients and 1,945 kinds of lab tests.
  - On average each clinical record contains 24.43 lab tests (1.26%) of all lab tests.
- We consider 3 complications:
  - hypertension (HTN), coronary heart disease (CHD), hyperlipidemia (HPL).
- SVM and traditional factor graph model suffer from the feature sparseness (-59.9% compared with our method by recall)
- PCA improve the performance. However, it separates sparse coding and classification into two processes.
- Our approach (SparseFGM) outperforms the baselines 20% by F1

**Association Analysis**

WBC in the urine typically is found in urinary tract infections which cause frequent voiding, which causes insomnia.

Hyperlipidemia (HPL) can be diagnosed more precisely, while depression (depr.) is usually recognized from psychological investigation instead of physiological lab tests.

**Input: a patient’s lab test results**

- Bilirubin example
- Routine urine analysis

**Output: diabetes complications**

- Heart
- Stroke
- 2-4 times increased risk
- Cause of amputation
- PAD
- Eye
- Retinopathy
- CRF
- Leading cause
- Diabetic retinopathy

**Model the joint distribution of a given set of lab tests X over complication labels Y as**

\[ p(y_n|x_n) = \prod_{k=1}^{K} \theta_{nk} \cdot \Omega_{x_nk} \]

The feature factor is defined as

\[ P(y_n|x_n) = \frac{1}{Z_1} \exp \left( \alpha \cdot f(x_n, y_n) \right) \]

The correlation factor is defined as

\[ P(y_n, y_{n'}) = \frac{1}{Z_2} \exp \left( \beta \cdot g(y_n, y_{n'}) \right) \]

The correlation factor is defined as

**Table 2: Performance of diabetes complication forecasting**

<table>
<thead>
<tr>
<th>Complication</th>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTN</td>
<td>SVM</td>
<td>0.3804</td>
<td>0.4789</td>
<td>0.4241</td>
</tr>
<tr>
<td></td>
<td>FGM</td>
<td>0.5666</td>
<td>0.4859</td>
<td>0.5075</td>
</tr>
<tr>
<td></td>
<td>FGM+PCA</td>
<td>0.5741</td>
<td>0.3284</td>
<td>0.4718</td>
</tr>
<tr>
<td></td>
<td>SparseFGM</td>
<td>0.4714</td>
<td>0.6319</td>
<td>0.5400</td>
</tr>
<tr>
<td>CHD</td>
<td>SVM</td>
<td>0.2132</td>
<td>0.0636</td>
<td>0.0090</td>
</tr>
<tr>
<td></td>
<td>FGM</td>
<td>0.6264</td>
<td>0.3169</td>
<td>0.2247</td>
</tr>
<tr>
<td></td>
<td>FGM+PCA</td>
<td>0.2425</td>
<td>0.3671</td>
<td>0.3832</td>
</tr>
<tr>
<td></td>
<td>SparseFGM</td>
<td>0.2522</td>
<td>0.7972</td>
<td>0.3832</td>
</tr>
<tr>
<td>HPL</td>
<td>SVM</td>
<td>0.2208</td>
<td>0.0460</td>
<td>0.0076</td>
</tr>
<tr>
<td></td>
<td>FGM</td>
<td>0.6557</td>
<td>0.0951</td>
<td>0.1084</td>
</tr>
<tr>
<td></td>
<td>FGM+PCA</td>
<td>0.2047</td>
<td>0.8035</td>
<td>0.3262</td>
</tr>
<tr>
<td></td>
<td>SparseFGM</td>
<td>0.2796</td>
<td>0.4896</td>
<td>0.4195</td>
</tr>
</tbody>
</table>

**Micro-level: association between 10 complications and 9 parameters of routine urine analysis discovered by the proposed mode.**

Contact: SherlockBourne@gmail.com
Code & Data: http://aminer.org/diabetes
Tsinghua University