Mining Information Dependency in Outpatient Encounters for Chronic Disease Care

Wen Sun, Weijia Shen, Xiang Li, Feng Cao, Yuan Ni, Haifeng Liu, Guotong Xie
IBM Research – China
Background

- Rapid increase of the prevalence of the chronic diseases in China
  - The prevalence of hypertension in the people over the age of 18 is 18.8% (more than 160 million)
  - There are more than 20 million diabetic patients in China, and nearly 20 million people with impaired fasting blood sugar level
- Insufficiency & inefficient use of healthcare resources
  - A primary care clinician in Shanghai manages over 100 patients of chronic disease
  - A patient typically visits multiple hospitals during chronic disease treatment

Source:
- IBM Institute for Business Value analysis.
Discover Clinical Information Dependency

Goal

- Improve health professionals’ working efficiency by discovering clinical information dependency and enabling proactive information provision

Clinical information dependency

- Can be used to collect related information for clinicians’ reference under particular working contexts
- Two types of (temporal) clinical information dependency
  - Co-occurrence
  - Sequential occurrence
- Related paper in MedInfo 2013
  - Using data mining techniques on discovering physician practice patterns regarding to medication prescription – an exploratory study
  - Improving physician practice Efficiency by learning lab test ordering pattern
Proposed Method

- Framework
Proposed Method

- Framework
Proposed Method

- Step 0: Outpatient encounter records
  - Over 10,000 type-2 diabetes patients of three hospitals in Shanghai

<table>
<thead>
<tr>
<th></th>
<th>Level-2 Hospital A</th>
<th>Comm. Hospital B</th>
<th>Comm. Hospital C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>122062</td>
<td>25859</td>
<td>26750</td>
<td>163723</td>
</tr>
<tr>
<td>Encounters</td>
<td>560235</td>
<td>217478</td>
<td>169125</td>
<td>946838</td>
</tr>
<tr>
<td>Diabetes Patients</td>
<td>7324</td>
<td>3551</td>
<td>3477</td>
<td>11946</td>
</tr>
<tr>
<td>Diabetes Patients’ Encounters</td>
<td>124209</td>
<td>44309</td>
<td>67177</td>
<td>235695</td>
</tr>
</tbody>
</table>
Proposed Method

- Step 1: Patient grouping
  - Makes the discovered information dependency more specific for each patient group
  - Group patients under certain criteria
    - Disease / complications, care providers, etc.
    - Similarity measures, e.g. Sun J. et al. “Supervised patient similarity measure of heterogeneous patient records”, ACM SIGKDD Explorations Newsletter, 2012

<table>
<thead>
<tr>
<th>Group ID</th>
<th>Group Description</th>
<th>Number of Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All the diabetes patients</td>
<td>11946</td>
</tr>
<tr>
<td>2</td>
<td>Diabetes patients with cardiovascular diseases</td>
<td>2310</td>
</tr>
<tr>
<td>3</td>
<td>Diabetes patients with cerebrovascular diseases</td>
<td>917</td>
</tr>
<tr>
<td>4</td>
<td>Diabetes patients with encounters in the level-2 hospital A</td>
<td>7324</td>
</tr>
<tr>
<td>5</td>
<td>Diabetes patients with encounters in the community hospital B</td>
<td>3551</td>
</tr>
<tr>
<td>6</td>
<td>Diabetes patients with encounters in the community hospital C</td>
<td>3477</td>
</tr>
</tbody>
</table>
Proposed Method

Step 2: Information Aggregation & Filtering

- Aggregate entries (medications, lab tests, etc.) based on categories
  - There are over 1,000 different diagnosis codes, over 1,500 different medications, and about 300 different lab test items in the encounter records
  - Diabetes medications: biguanides, insulin secretagogues, insulin sensitizers, a-glycosidase inhibitors, DPP-4 inhibitors, and insulin injections.
  - Lab tests: tens of groups
  - Diabetes-related diagnosis: diabetes retinopathy complications, diabetes nephropathy complications, etc.

- Aggregate records using on time windows

- Filtering confounding factors
  - Upper respiratory infection from diabetes;
  - Aspirin and sodium chloride in medications
Proposed Method

- **Step 3: Information Dependency Mining**
  - Co-occurrence: frequent itemset mining (Szathmary L. et al. 2007)
  
  \[
  T(T_0) = \left\{ \begin{array}{l}
  (D'_{00}, M'_{00}, L'_{00}), (D'_{01}, M'_{01}, L'_{01}), \ldots, (D'_{ON}, M'_{ON}, L'_{ON}) \\
  (D'_{10}, M'_{10}, L'_{10}), (D'_{11}, M'_{11}, L'_{11}), \ldots, (D'_{1N}, M'_{1N}, L'_{1N}) \\
  \ldots \ldots \ldots \ldots \\
  (D'_{K0}, M'_{K0}, L'_{K0}), (D'_{K1}, M'_{K1}, L'_{K1}), \ldots, (D'_{KN}, M'_{KN}, L'_{KN}) 
  \end{array} \right. 
  \]

  - Sequential dependency: Frequent sequential pattern mining (Fournier-Viger et al. 2008)

  \[
  S(T_0, T_1) = \left\{ \begin{array}{l}
  \ldots, (D'_{00}, M'_{00}, L'_{00}; D''_{00}, M''_{00}, L''_{00}), \ldots \\
  \ldots, (D'_{10}, M'_{10}, L'_{10}; D''_{10}, M''_{10}, L''_{10}), \ldots \\
  \ldots \ldots \ldots \ldots \\
  \ldots, (D'_{K0}, M'_{K0}, L'_{K0}; D''_{K0}, M''_{K0}, L''_{K0}), \ldots \\
  \ldots \ldots \ldots \ldots \\
  \ldots, (D'_{K1}, M'_{K1}, L'_{K1}; D''_{K1}, M''_{K1}, L''_{K1}), \ldots \\
  \ldots \ldots \ldots \ldots \\
  \ldots, (D'_{K_N}, M'_{K_N}, L'_{K_N}; D''_{K_N}, M''_{K_N}, L''_{K_N}), \ldots 
  \end{array} \right. 
  \]
## Results

- Number of the discovered patterns of co-occurrence (Co.) and sequential (Seq.) information dependency under different time window settings (0, 7, 15, 30 days)

<table>
<thead>
<tr>
<th>Group ID</th>
<th>$T_0=T_1=0$</th>
<th>$T_0=T_1=7$</th>
<th>$T_0=T_1=15$</th>
<th>$T_0=T_1=30$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>146</td>
<td>-</td>
<td>162</td>
<td>58</td>
</tr>
<tr>
<td>2</td>
<td>73</td>
<td>-</td>
<td>93</td>
<td>41</td>
</tr>
<tr>
<td>3</td>
<td>61</td>
<td>-</td>
<td>87</td>
<td>39</td>
</tr>
<tr>
<td>4</td>
<td>101</td>
<td>-</td>
<td>135</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>82</td>
<td>-</td>
<td>110</td>
<td>38</td>
</tr>
<tr>
<td>6</td>
<td>96</td>
<td>-</td>
<td>114</td>
<td>42</td>
</tr>
</tbody>
</table>
Results

- Frequent co-occurrence diabetes medications
Results

- Top 3 sequential patterns for taking different types of diabetes medications

<table>
<thead>
<tr>
<th>Rank</th>
<th>Mined Sequential Pattern</th>
<th>Average Transition Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(biguanides ; insulin secretagogues)</td>
<td>9.3</td>
</tr>
<tr>
<td>2</td>
<td>(insulin secretagogues ; insulin injections)</td>
<td>17.1</td>
</tr>
<tr>
<td>3</td>
<td>(insulin secretagogues ; biguanides)</td>
<td>11.5</td>
</tr>
</tbody>
</table>

Group 1, 30-day time window
Conclusion and Discussion

- We proposed an effective approach to discover information dependency in real-world outpatient encounter data for chronic disease care, aiming to facilitate information provision and sharing in the care coordination of multiple care providers.

- The discovered information dependency relations can be specific to some patient cohort or some diseases.

- The data-driven approach is complementary to clinical guidelines.

- The integration of such information dependency to clinician’s practice remains to be studied.
Thank You

- English
- Simplified Chinese
- Italian
- Tamil

감사합니다
多謝
ขอบคุณ
多謝
شكرا
多謝
Ευχαριστώ
Grazie
Obrigado
Danke
Merci
धन्यवाद

Korean
Traditional Chinese
Thai
Russian
Arabic
Greek
Brazilian Portuguese
German
French
Hindi
Japanese
Tamil

Smarter Healthcare