Creating Interpretable Collaborative Patterns for Auditing

You Chen
Department of Biomedical Informatics, Vanderbilt University
You.chen@vanderbilt.edu
http://hiplab.org/~ychen
6-Nearest Neighbor Network-Vanderbilt Medical Center (1 day of accesses)

(Chen, Nyemba, & Malin – IEEE TDSC 2012)
Deviation of Accesses on CADS Vanderbilt Dataset

Audit These users!
CADS on Northwestern Dataset

Audit These users!
But Relationships Decay...
(Malin, Nyemba, & Paulett – JBI 2011)

- EMR <user, user> relationships
- EMR <department, department> relationships
Department Level

Auditing Medical Record Accesses through Department Interactions
Hospital Departmental Relations Can Be Inferred
(Chen, Nyemba, & Malin - AMIA 2012)

- Probability department $d_i$ accesses a patient’s record, given department $d_j$ accessed the record.

<table>
<thead>
<tr>
<th>Department ($d_i$)</th>
<th>Department ($d_j$)</th>
<th>Min Certainty</th>
<th>Max Certainty</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intradepartmental Relations</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4East OB/GYN</td>
<td>4East OB/GYN</td>
<td>0.74319</td>
<td>0.7669</td>
</tr>
<tr>
<td>Adult Emergency Medicine</td>
<td>Adult Emergency Medicine</td>
<td>0.74024</td>
<td>0.78453</td>
</tr>
<tr>
<td>Cancer Infusion Center</td>
<td>Cancer Infusion Center</td>
<td>0.73171</td>
<td>0.844</td>
</tr>
<tr>
<td>8N Inpatient Medicine</td>
<td>8N Inpatient Medicine</td>
<td>0.7197</td>
<td>0.80909</td>
</tr>
<tr>
<td>Newborn Nursery</td>
<td>Newborn Nursery</td>
<td>0.70406</td>
<td>0.72727</td>
</tr>
<tr>
<td><strong>Interdepartmental Relations</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOT Radiology</td>
<td>Orthopaedics</td>
<td>0.99621</td>
<td>1</td>
</tr>
<tr>
<td>Nursing Education and Development</td>
<td>Medical Information Services</td>
<td>0.95833</td>
<td>1</td>
</tr>
<tr>
<td>Main OR - Trauma/Renal</td>
<td>Medical Information Services</td>
<td>0.94444</td>
<td>1</td>
</tr>
<tr>
<td>Life Flight Event Medicine</td>
<td>Emergency Medicine</td>
<td>0.90805</td>
<td>1</td>
</tr>
<tr>
<td>Emergency Medicine Admin</td>
<td>Adult Emergency Medicine</td>
<td>0.91489</td>
<td>0.94186</td>
</tr>
</tbody>
</table>
Organization Level-Department

Tripartite graph of departments, users and patients

Bipartite graph of departments and patients

Interaction network of departments

Local view for $p_6$

Global view
Certainty to Model Relationship Among Departments

Cert(Emergency medicine (d₁)→Lifeflight event medicine (d₃)) = 4/7
Lifeflight event medicine (d₃)→ Cert(Emergency medicine (d₁)) = 4/4
(a) A bipartite graph of departments and subjects

(b) A social graph of departments based on common subjects accessed

(c) A confidence social network of departments
Evolution of A Global Network Over the Time
The changes become smaller over time (centralization: green > blue > red)

Degree of relations between departments changes little over time

>82.5% of the change resides in [-0.25, 0.25]
Using reciprocity to characterize the mutual interaction between all pairs of departments in the global network

Reciprocity = 1

Inpatient Admin -> VUH Admitting 0.75

VUH Admitting -> Inpatient Admin 0.12
Although the relations of the network are very unbalanced, the unbalance is stable over time.

<table>
<thead>
<tr>
<th>Time</th>
<th>Week 1</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reciprocity</td>
<td>0.267</td>
<td>0.2814</td>
<td>0.2858</td>
<td>0.2871</td>
</tr>
</tbody>
</table>

\[
\text{Reciprocity} = \frac{\sum_{d_i, d_j \in D, i \neq j} \left| (\text{Cert}(d_i \rightarrow d_j) - a) \times (\text{Cert}(d_j \rightarrow d_i) - a) \right|}{\sum_{d_i, d_j \in D, i \neq j} (\text{Cert}(d_i \rightarrow d_j) - a)^2}
\]

\[
a = \frac{\sum_{d_i, d_j \in D, i \neq j} \text{Cert}(d_i \rightarrow d_j)}{2 \times |E|}
\]
Evolution of Local Network Relations Can be Used Detect “Strange” Behavior

Each point in the $P_{\text{start}}$ corresponds to a local network

A Local network for $p_6$
Over 80% of local networks whose size is less than 5

A Local network or a patient
Over 80% of local networks has number of departments change less than 2
Most Patients Network Suggest They Are “Normal”

Audit These Patients!
Approximately 99% of patients are normal because they have a change of reciprocity <0.1
p2 has -0.93 change of local network score and -0.79 change of local reciprocity from the 1st to the 2nd week.

**Breast Center, [Anonymized Street Location], Care/Eskind Diab Acces, Disease Management Service, Eskind Diabetes - Adult, Free Stipends, Internal Medicine, VIM, VMG Physician Billing Services, Vanderbilt Home Care Primary**
Findings

• We hypothesized an HCO would exhibit strong stability → confirmed by our experiments
• The changes in the score of local networks do not justify the claim that the patient has been intruded upon, but may provide a reason for an investigation that incorporates more nuanced domain knowledge
But Do You Believe the Data?
Survey Population

(Chen, Lorenzi, Nyemba, Schildcrout & Malin – IJMI 2014)

- Vanderbilt University Medical Center areas

<table>
<thead>
<tr>
<th>Anesthesiology</th>
<th>Psychiatry</th>
<th>Coding &amp; Charge Entry</th>
<th>Medical Information Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 members</td>
<td>10 members</td>
<td>10 members</td>
<td>10 members</td>
</tr>
</tbody>
</table>

**Invited**

- 10 members

**Viewed**

- 8
- 9
- 7
- 10

**Completed**

- 4
- 6
- 7
- 9

34 respondents did the survey and 26 of them are valuable
Survey Questions

- Departmental interactions
- Conditional probabilities of accessing a record (conditioned on the HCO area)
- “Given someone from Coding & Charge Entry accessed the record, what's the chance someone from the following Area accessed the record?

**Emergency Medicine**

1. Not at all Likely
2. Slightly Likely
3. Moderately Likely
4. Very Likely
5. Completely Likely
Coding & Charge Entry Interactions
(one week, ~620 points)
Survey Questions

Coding & Charge Entry

High

Low

Psychiatry

Anesthesiology

Medical Information Services

20 rules
Survey Questions

Coding & Charge Entry
- High
- Low

Psychiatry
- High
- Low

Anesthesiology
- High
- Low

Medical Information Services
- High
- Low

20 rules
20 rules
20 rules
20 rules
CODE RESPONDENT(7)

MIS Rules

PSY Rules

CODE Rules

ANE Rules
Hypothesis

• 1) Employees can distinguish between high, and low likelihood rules for all HCO areas
• 2) Employees can distinguish between high and low likelihood rules for their own HCO area
• 3) Employees can distinguish between high, and low likelihood rules in their own HCO area better than they can in other HCO areas
One respondent has 8 observations
The total number of observations is $8 \times 26 = 208$

<table>
<thead>
<tr>
<th>Respondent (ID)</th>
<th>Respondent Type</th>
<th>Rule Type</th>
<th>Rule Class</th>
<th>Average Score of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MIS</td>
<td>ANE</td>
<td>High</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>MIS</td>
<td>ANE</td>
<td>Low</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>MIS</td>
<td>CODE</td>
<td>High</td>
<td>3.3</td>
</tr>
<tr>
<td>1</td>
<td>MIS</td>
<td>CODE</td>
<td>Low</td>
<td>2.1</td>
</tr>
<tr>
<td>1</td>
<td>MIS</td>
<td>MIS</td>
<td>High</td>
<td>3.1111</td>
</tr>
<tr>
<td>1</td>
<td>MIS</td>
<td>MIS</td>
<td>Low</td>
<td>2.125</td>
</tr>
<tr>
<td>1</td>
<td>MIS</td>
<td>PSY</td>
<td>High</td>
<td>2.9</td>
</tr>
<tr>
<td>1</td>
<td>MIS</td>
<td>PSY</td>
<td>Low</td>
<td>2.05</td>
</tr>
</tbody>
</table>
Hypothesis Test 1 – Rules of All HCO Areas:

- Anesthesiologists
- Psychiatrists
- Coders
- MIS Specialists

High
Low
High
Low
High
Low
High
Low
Hypothesis Test1 – Rules of All HCO Areas:
One-sided t-test, 95% confidence

Confirmed!

Anesthesiologists

Confirmed!

Psychiatrists

Confirmed!

Coders

Confirmed!

Medical Information Specialists

High

Low

High

Low

Low

High

Low

High

Low
Linear Mixed Effects Model

Rule class: high or low
Respondent type: MIS, CODE, PSY, ANE
Rules type: MIS, CODE, PSY, ANE

\[ \text{lemr} \left( \text{aveScore} \sim h \times p + r + (1 \mid id), \text{data} \right) \]

\[ Y = \beta_0 + \beta_1(h = 1) + \beta_2(p = \text{code}) + \beta_3(p = \text{psy}) + \beta_4(p = \text{ane}) \]
\[ + \beta_5(r = \text{code}) + \beta_6(r = \text{psy}) + \beta_7(r = \text{ane}) \]
\[ + \beta_8(p = \text{code})(h = 1) + \beta_9(p = \text{psy})(h = 1) + \beta_{10}(p = \text{ane})(h = 1) \]

How MIS respondents distinguish high and low likelihood of rules for all HCO areas

\[ E(y \mid p=\text{mis}, h=0, r=\text{ALL}) = \beta_0 + \beta_5 + \beta_6 + \beta_7; \]
\[ E(y \mid p=\text{mis}, h=1, r=\text{ALL}) = \beta_0 + \beta_1 + \beta_5 + \beta_6 + \beta_7; \]
\[ E(y \mid p=\text{mis}, h=1, r=\text{ALL}) - E(y \mid p=\text{mis}, h=0, r=\text{ALL}) = \beta_1 \]

How about CODE, PSY and ANE?
Certain respondents are inclined to assign large likelihoods (upper right section of the plot),

while others are included to assign small likelihoods (the lower left section of the plot).
<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\beta$ values</th>
<th>description</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0.351557</td>
<td>MIS respondents distinguish high and low likelihood rules for all HCO areas</td>
<td>1.91*10^{-9}</td>
</tr>
<tr>
<td>$\beta_1+\beta_8$</td>
<td>0.521492</td>
<td>CODE respondents distinguish high and low likelihood rules for all HCO areas</td>
<td>1.11*10^{-6}</td>
</tr>
<tr>
<td>$\beta_1+\beta_9$</td>
<td>0.677858</td>
<td>PSY respondents distinguish high and low likelihood rules for all HCO areas</td>
<td>9.33*10^{-8}</td>
</tr>
<tr>
<td>$\beta_1+\beta_{10}$</td>
<td>0.691166</td>
<td>ANE respondents distinguish high and low likelihood rules for all HCO areas</td>
<td>1.22*10^{-8}</td>
</tr>
</tbody>
</table>

Respondents from four areas can distinguishing between high and low likelihood rules for **all HCO areas**
Hypothesis Test 2– Self Assessment:
Linear Mixed Effects Model
One-sided t-test, 95% confidence

Anesthesiologists

Psychiatrists

Coders

Anesthesiologists

Confirmed!

Confirmed!

Confirmed!

Confirmed!
Hypothesis Test 3—Bias Toward Own Rules
Linear Mixed Effects Model
One-sided t-test, 95% confidence

Respondents from HCO area ANE are better at distinguishing between high and low likelihood rules classes associated with area ANE than area PSY, CODE and MIS.
Hypothesis Test 3 – Bias Toward Own Rules

Linear Mixed Effects Model
One-sided t-test, 95% confidence
Conclusions

• Healthcare organization employees generally understand what goes on around them...
  ... and for other sections of the organization as well!

• Automated healthcare organizational modeling may be possible.

• Anomalies detection through collaborative patterns may be reliable!
Q&A
Thanks!