Bringing Big Data to Personalized Healthcare: A Patient-Centered Framework

Nitesh Chawla, PhD
Frank Freimann Collegiate Associate Professor
Computer Science and Engineering
Director, iCeNSA
National Health Expenditures, 2010

Total - $2.594 trillion


http://icensana.nd.edu

Nitesh Chawla, PhD (c)
Patient empowerment. Electronic health records will help empower patients to take a more active role in their health and in the health of their families.

“Health care has been evolving away from a ‘disease-centered model’ and toward a ‘patient-centered model.’ In the older, disease-centered model, physicians make almost all treatment decisions based largely on clinical experience and data from various medical tests. In a patient-centered model, patients become active participants in their own care and receive services designed to focus on their individual needs and preferences, in addition to advice and counsel from health professionals.” AHRQ.GOV
Two thousand years ago...

“It is far more important to know what person the disease has than what disease the person has,”

Hippocrates
What are **my** disease risks? *A Personalized Approach*

“Determine **individual** risk of developing specific diseases, detect the disease’s **earliest** onset, and **prevent** or intervene early enough to provide maximum benefit”
Empowering the patient and physician with the inferences drawn from millions of other patients

CARE: Collaborative Assessment and Recommendation Engine
Patent No. 8,504,343
“Data is a vital raw material”
Partnership is THE Enabler

- 13 Million ICD-9-CM data (Medicare)
  - In-patient
  - 32 Million visits over 4 years
- 500,000 out-patient data from a regional health system
  - Spread over 12 years

- ICD-9-CM Diagnosis 402
  - Hypertensive heart disease
    - 402 is a non-specific code that cannot be used to specify a diagnosis

- ICD-9-CM Diagnosis 402.0
  - Malignant hypertensive heart disease
    - 402.0 is a non-specific code that cannot be used to specify a diagnosis
    - 402.0 contains 3 index entries

- ICD-9-CM Diagnosis 402.01
  - Malignant hypertensive heart disease with heart failure
    - 402.01 is a specific code that can be used to specify a diagnosis
    - 402.01 contains 6 index entries

http://icensa.nd.edu
\[ P_{a,j} = \bar{v}_j + K_a \sum_{i \in I_j} w(a, i)(v_{i,j} - \bar{v}_i) \]

- \( i \) - training patient
- \( j \) - disease
- \( a \) - the active patient
- \( I \) - set of all training patients
- \( I_j \) - set of all training patients with disease \( j \)
- \( w(a, i) \) - the similarity between user \( a \) and user \( i \)
ICARE – an Iterative Version

- Problem: Even with IFVS, common diseases overwhelm rare diseases, but tend to be more obvious and less interesting.

- Solution: Disease-focused ensembles

\[
\max_{c \in G} \left( \bar{v}_{j,c} + \kappa (1 - \bar{v}_{j,c}) \sum_{i \in I_{j,c}} w(a, i) \right)
\]
Time Sensitive iCARE

- Patients should be more similar as disease progression is similar, not just diagnoses
- Many medical events
- Want to further limit data to the specific period of interest in each patient’s life
Time Sensitive iCARE

• Goal: Find best match (highest VS) subset of consecutive visits from training patient’s record

• Special version of maximum subsequence problem
  • Linear scan solution
Experimental Setup

- Predictions are only for future diseases
- Patient must have at least 5 visits

<table>
<thead>
<tr>
<th>Round</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Visit 1</td>
<td>Visits 2-5</td>
</tr>
<tr>
<td>2</td>
<td>Visits 1-2</td>
<td>Visits 3-5</td>
</tr>
<tr>
<td>3</td>
<td>Visits 1-3</td>
<td>Visits 4-5</td>
</tr>
<tr>
<td>4</td>
<td>Visit 1-4</td>
<td>Visit 5</td>
</tr>
</tbody>
</table>
# iCARE Results

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>3-digit ICARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td>0.321</td>
<td>0.513</td>
</tr>
<tr>
<td>Average Rank</td>
<td>7.326</td>
<td>5.668</td>
</tr>
</tbody>
</table>

Nitesh Chawla, PhD (c)

http://icensa.nd.edu
Load Diseases

Load Patients

Best Match Calculation

All patients

Iterate Over

Avg Num of Unique Diseases

Total Patients (log scale)

Avg Num of Unique Diseases

5 10 25 50 100 250 500 1000 2000

Nitesh Chawla, PhD (c)
Leverage for:

1) personalized and prospective healthcare
2) improved chronic disease management
3) reduction in medical errors and ADEs
4) reduced healthcare costs
5) best practices and efficiencies in delivery of healthcare
6) educational opportunities
7) exciting research opportunities and grand challenge problems for data mining and computer science
Technology solve: Empower, personalize and sustain

Initial Wellness Assessment

What are my risks?

Disease Management  Nutrition/Exercise Program

What should I do?

Improved health and well-being

Sustain: Wellness

http://icensa.nd.edu


Thank you

- nchawla@nd.edu
- http://icensa.nd.edu